10/825,622

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-306205

(43)Date of publication of application : 05.11.1999

(51)Int.Cl. G06F 17/30

(21)Application number : 10-129485

(22)Date of filing : 23.04.1998

(71)Applicant : NEC CORP

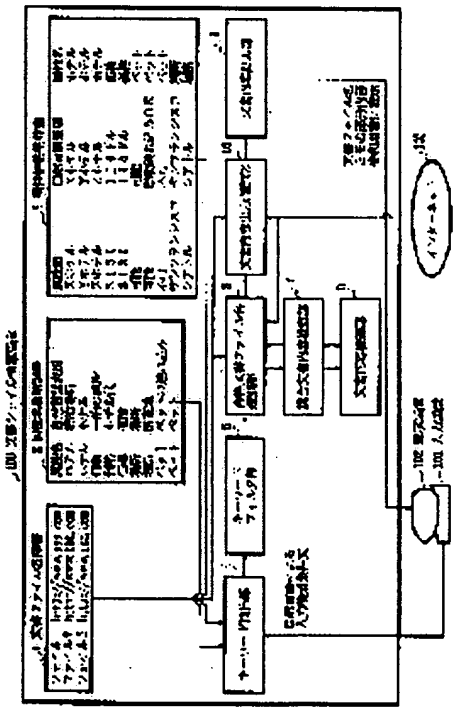
(72)Inventor : SHIMAZU HIDEO

(54) DOCUMENT FILE RETRIEVAL DEVICE AND MACHINE READABLE RECORDING MEDIUM RECORDING PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To realize a retrieval inquiry about a WWW home page by a natural language.

SOLUTION: A WWW home page being a retrieval object document file is described in an XML. When a retrieval condition composition is inputted, a keyword extraction part 4 converts a natural language expression expressing an attribute name into an attribute name index including the attribute name and also converts the natural language expression expressing the attribute value into an attribute value index including a pair of the said attribute name and attribute value. A keyword filter part 5 deletes the attribute name index existing at a place where the attribute name and the attribute value of the same attribute exist adjacent to each other in a converted index string. A document contents check part 6 checks whether or not a tag corresponding to pairs of attribute name and value of the all attribute value index exists in the retrieval object document file. If the said tag exists, a document contents output part 9 retrieves and outputs the attribute value of the tag having the relevant attribute name of the attribute name index that is included in the converted index string.



LEGAL STATUS

[Date of request for examination] 23.04.1998

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3191762
[Date of registration]	25.05.2001
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号

特開平11－306205

(43)公開日　平成11年(1999)11月５日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ０６Ｆ　17/30		Ｇ０６Ｆ　15/403	３３０Ｃ
		15/40	３７０Ａ

審査請求　有　　請求項の数７　ＦＤ　（全 21 頁）

(21)出願番号　特願平10－129485

(22)出願日　平成10年(1998)４月23日

(71)出願人　000004237
日本電気株式会社
東京都港区芝五丁目７番１号

(72)発明者　島津　秀雄
東京都港区芝五丁目７番１号　日本電気株式会社内

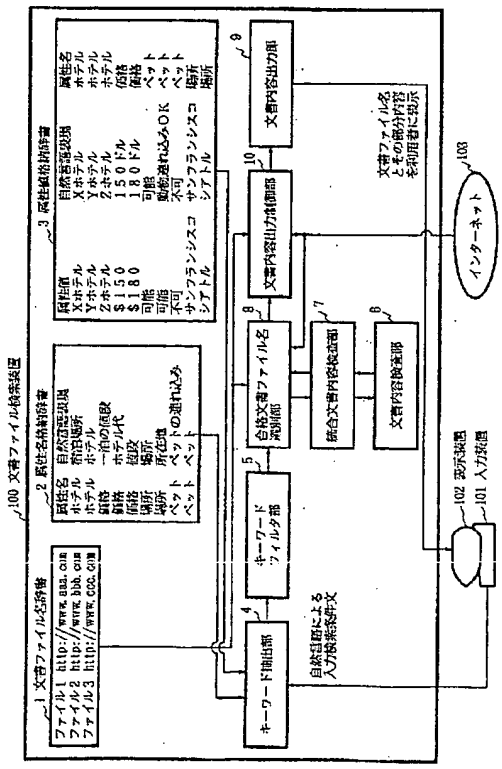
(74)代理人　弁理士　境　廣巳

(54)【発明の名称】　文書ファイル検索装置及びプログラムを記録した機械読み取り可能な記録媒体

(57)【要約】

【課題】　WWW のホームページに対する自然言語による検索問い合わせを実現する。

【解決手段】　検索対象文書ファイルであるWWW のホームページをXML で記述する。検索条件文が入力されるとキーワード抽出部４は、属性名を表現する自然言語表現はその属性名を含む属性名インデックスに、属性値を表現する自然言語表現はその属性値と属性名との対を含む属性値インデックスに変換する。キーワードフィルタ部５は、変換後のインデックス列中で、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する箇所の属性名インデックスを削除する。文書内容検査部６は、検索対象文書ファイル中に、変換後のインデックス列中の全ての属性値インデックスの属性名と属性値との対に対応するタグが存在するか否かを調べ、存在する場合、文書内容出力部９が変換後のインデックス列中の属性名インデックスの属性名を持つタグの属性値を検索して出力する。



【特許請求の範囲】

【請求項１】　属性の属性名とその属性の属性値との対を内蔵する文書ファイルを検索対象文書ファイルとし、検索対象文書ファイルから、利用者が自然言語で指定した検索条件に適合する部分を検索する文書ファイル検索装置において、

自然言語で表現した検索要求文を先頭から順に探査し、属性名を表現する自然言語表現に対してはその属性名を属性名インデックスとして出力し、属性値を表現する自然言語表現に対してはその属性値と属性名との対を属性値インデックスとして出力することを順次行うキーワード抽出部と、

前記キーワード抽出部の出力を入力して先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する場合のみ、前記属性名インデックスを削除し、それ以外の部分はそのまま出力するキーワードフィルタ部と、

検索対象文書ファイル中に、前記キーワードフィルタ部から出力された全ての属性値インデックスの属性名と属性値との対が内蔵されているか否かを調べ、内蔵されている場合、前記キーワードフィルタ部から出力された属性名インデックスの属性名に対応する属性値を検索対象文書ファイルから検索して出力する検索手段とを備えた文書ファイル検索装置。

【請求項２】　検索対象文書ファイル中に存在する属性名について、属性名とその属性名を表現する自然言語表現との対を格納しておく属性名格納辞書と、

検索対象文書ファイル中に存在する属性値について、属性値とその属性値に対応する属性名とその属性値を表現する自然言語表現との３つ組を格納しておく属性値格納辞書とを備え、

前記キーワード抽出部は、自然言語で表現した検索要求文を先頭から順に探査し、属性名格納辞書を参照して、属性名を表現する自然言語表現が含まれていたら、その自然言語表現と対である属性名を属性名インデックスとして出力し、属性値格納辞書を参照して、属性値を表現する自然言語表現が含まれていたら、その自然言語表現と３つ組である属性値と属性名との対の集合を属性値インデックスとして出力する構成を有することを特徴とする請求項１記載の文書ファイル検索装置。

【請求項３】　文書中に書かれた意味を表現する属性名のついたタグとその属性の属性値との対を内蔵する文書ファイルを検索対象文書ファイルとし、検索対象文書ファイルから、利用者が自然言語で指定した検索条件に適合する部分を検索する文書ファイル検索装置において、

検索対象文書ファイル中に存在する属性名について、属性名とその属性名を表現する自然言語表現との対の集合を格納しておく属性名格納辞書と、

検索対象文書ファイル中に存在する属性値について、属性値とその属性値に対応する属性名とその属性値を表現する自然言語表現との３つ組の集合を格納しておく属性値格納辞書と、

自然言語で表現した検索要求文を先頭から順に探査し、属性名格納辞書を参照して、属性名を表現する自然言語表現が含まれていたら、その自然言語表現と対である属性名を属性名インデックスとして出力し、属性値格納辞書を参照して、属性値を表現する自然言語表現が含まれていたら、その自然言語表現と３つ組である属性値と属性名との対を属性値インデックスとして出力するキーワード抽出部と、

キーワード抽出部の出力を入力し、先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する場合のみ、前記属性名インデックスを削除し、それ以外の部分はそのまま出力するキーワードフィルタ部と、

検索対象文書ファイル中に、前記キーワードフィルタ部から出力された全ての属性値インデックスの属性名と属性値との対に対応するタグの対が内蔵されているか否かを調べ、内蔵されている場合、前記キーワードフィルタ部から出力された属性名インデックスの属性名を持つタグの属性値を検索対象文書ファイルから検索して出力する検索手段とを備えた文書ファイル検索装置。

【請求項４】　文書中に書かれた意味を表現する属性名のついたタグとその属性の値との対を複数個内蔵する文書ファイルの集合から、利用者が自然言語で指定した検索条件を満足する文書ファイルを選択してその適合する部分を表示する文書ファイル検索装置において、

検索対象となるすべての文書ファイルの名前と存在位置とを格納する文書ファイル名辞書と、

検索対象となる文書ファイル中に存在する属性名について、属性名とその属性名を表現する自然言語表現との対の集合を格納しておく属性名格納辞書と、

検索対象となる文書ファイル中に存在する属性値について、属性値とその属性値に対応する属性名とその属性値を表現する自然言語表現との３つ組の集合を格納しておく属性値格納辞書と、

利用者が、自然言語で表現した検索要求文を入力すると、前記入力文を先頭から順に探査し、属性名格納辞書を参照して、属性名を表現する自然言語表現が含まれていたら、その自然言語表現と対である属性名を属性名インデックスとして出力し、属性値格納辞書を参照して、属性値を表現する自然言語表現が含まれていたら、その自然言語表現と３つ組である属性値と属性名との対の集合を属性値インデックスとして出力することを順次行うキーワード抽出部と、

キーワード抽出部の出力を入力し、先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣通しに存在する場合のみ、前記属性名インデックスを削除し、それ以外の部分はそのまま出力するキーワードフィルタ部と、

文書ファイルの内容と属性値インデックスとを入力すると、前記文書ファイルの内容中に、前記属性値インデックス中の属性名を含むタグが存在するかどうか調べ、存在する場合は、そのタグと対で存在する属性値を取り出し、その値が前記属性値インデックス中の属性値と等しいかどうか調べ、等しい場合は、合格の出力をし、そうでない場合は不合格の出力をする文書内容検査部と、
文書ファイルの内容と１つ以上の属性値インデックスとを入力すると、前記属性値インデックスから１つずつ取り出し、前記文書ファイルの内容と前記取り出した属性値インデックスとを１つずつ文書内容検査部に渡していき、すべての属性値インデックスに対してその出力が合格のときは、合格を出力し、そうでないときは不合格を出力する統合文書内容検査部と、
前記文書ファイル名辞書を参照して、１つずつ文書ファイルの内容を取り出し、前記文書の内容とキーワードフィルタ部の出力のうちの属性値インデックスの部分とを統合文書内容検査部に渡し、前記統合文書内容検査部の出力を受け取ることを前記１つずつ取り出した文書ファイルのすべてに対して行い、前記出力が合格の文書ファイルの名前のみを出力する合格文書ファイル名選別部と、
文書ファイル名と前記文書ファイル名の内容とキーワードフィルタ部の出力である属性名インデックスとを入力すると、前記属性名インデックスのうちの１つを取り出し、与えられた前記文書ファイルの内容中に、前記取り出した属性名を含むタグが存在するかどうか調べ、存在する場合は、その属性名のタグの値と前記入力した文書ファイル名とを利用者に表示し、存在しない場合には何も出力しないことを、前記入力した属性名インデックスのそれぞれに対して行う文書内容出力部と、
前記合格文書ファイル名選別部の出力である文書ファイル名の集合を入力し、文書ファイル名格納辞書を参照して、前記入力した文書ファイル名の集合の要素を１つずつ取り出し、文書内容出力部に渡すことを、前記入力中の文書ファイル名のすべてに対して行うことを繰り返す文書内容出力制御部とを備えることを特徴とする文書ファイル検索装置。

【請求項５】　属性の属性名とその属性の属性値との対を内蔵する文書ファイルを検索対象文書ファイルとし、検索対象文書ファイルから、利用者が自然言語で指定した検索条件に適合する部分を検索する文書ファイル検索装置を構成するコンピュータを、
自然言語で表現した検索要求文を先頭から順に探査し、属性名を表現する自然言語表現に対してはその属性名を属性名インデックスとして出力し、属性値を表現する自然言語表現に対してはその属性値と属性名との対を属性値インデックスとして出力することを順次行うキーワード抽出部、
前記キーワード抽出部の出力を入力して先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する場合のみ、前記属性名インデックスを削除し、それ以外の部分はそのまま出力するキーワードフィルタ部、
検索対象文書ファイル中に、前記キーワードフィルタ部から出力された全ての属性値インデックスの属性名と属性値との対が内蔵されているか否かを調べ、内蔵されている場合、前記キーワードフィルタ部から出力された属性名インデックスの属性名に対応する属性値を検索対象文書ファイルから検索して出力する検索手段、
として機能させるプログラムを記録した機械読み取り可能な記録媒体。

【請求項６】　文書中に書かれた意味を表現する属性名のついたタグとその属性の属性値との対を内蔵する文書ファイルを検索対象文書ファイルとし、検索対象文書ファイルから、利用者が自然言語で指定した検索条件に適合する部分を検索する文書ファイル検索装置を構成するコンピュータを、
検索対象文書ファイル中に存在する属性名について、属性名とその属性名を表現する自然言語表現との対の集合を格納しておく属性名格納辞書、
検索対象文書ファイル中に存在する属性値について、属性値とその属性値に対応する属性名とその属性値を表現する自然言語表現との３つ組の集合を格納しておく属性値格納辞書、
自然言語で表現した検索要求文を先頭から順に探査し、属性名格納辞書を参照して、属性名を表現する自然言語表現が含まれていたら、その自然言語表現と対である属性名を属性名インデックスとして出力し、属性値格納辞書を参照して、属性値を表現する自然言語表現が含まれていたら、その自然言語表現と３つ組である属性値と属性名との対を属性値インデックスとして出力するキーワード抽出部、
キーワード抽出部の出力を入力し、先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する場合のみ、前記属性名インデックスを削除し、それ以外の部分はそのまま出力するキーワードフィルタ部、
検索対象文書ファイル中に、前記キーワードフィルタ部から出力された全ての属性値インデックスの属性名と属性値との対に対応するタグの対が内蔵されているか否かを調べ、内蔵されている場合、前記キーワードフィルタ部から出力された属性名インデックスの属性名を持つタグの属性値を検索対象文書ファイルから検索して出力する検索手段、
として機能させるプログラムを記録した機械読み取り可能な記録媒体。

【請求項７】　文書中に書かれた意味を表現する属性名のついたタグとその属性の値との対を複数個内蔵する文書ファイルの集合から、利用者が自然言語で指定した検

索条件を満足する文書ファイルを選択してその適合する部分を表示する文書ファイル検索装置を構成するコンピュータを、
検索対象となるすべての文書ファイルの名前と存在位置とを格納する文書ファイル名辞書、
検索対象となる文書ファイル中に存在する属性名について、属性名とその属性名を表現する自然言語表現との対の集合を格納しておく属性名格納辞書、
検索対象となる文書ファイル中に存在する属性値について、属性値とその属性値に対応する属性名とその属性値を表現する自然言語表現との３つ組の集合を格納しておく属性値格納辞書、
利用者が、自然言語で表現した検索要求文を入力すると、前記入力文を先頭から順に探査し、属性名格納辞書を参照して、属性名を表現する自然言語表現が含まれていたら、その自然言語表現と対である属性名を属性名インデックスとして出力し、属性値格納辞書を参照して、属性値を表現する自然言語表現が含まれていたら、その自然言語表現と３つ組である属性値と属性名との対の集合を属性値インデックスとして出力することを順次行うキーワード抽出部、
キーワード抽出部の出力を入力し、先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣通しに存在する場合のみ、前記属性名インデックスを削除し、それ以外の部分はそのまま出力するキーワードフィルタ部、
文書ファイルの内容と属性値インデックスとを入力すると、前記文書ファイルの内容中に、前記属性値インデックス中の属性名を含むタグが存在するかどうか調べ、存在する場合は、そのタグと対で存在する属性値を取り出し、その値が前記属性値インデックス中の属性値と等しいかどうか調べ、等しい場合は、合格の出力をし、そうでない場合は不合格の出力をする文書内容検査部、
文書ファイルの内容と１つ以上の属性値インデックスを入力すると、前記属性値インデックスから１つずつ取り出し、前記文書ファイルの内容と前記取り出した属性値インデックスとを１つずつ文書内容検査部に渡していき、すべての属性値インデックスに対してその出力が合格のときは、合格を出力し、そうでないときは不合格を出力する統合文書内容検査部、
前記文書ファイル名辞書を参照して、１つずつ文書ファイルの内容を取り出し、前記文書の内容とキーワードフィルタ部の出力のうちの属性値インデックスの部分とを統合文書内容検査部に渡し、前記統合文書内容検査部の出力を受け取ることを前記１つずつ取り出した文書ファイルのすべてに対して行い、前記出力が合格の文書ファイルの名前のみを出力する合格文書ファイル名選別部、
文書ファイル名と前記文書ファイル名の内容とキーワードフィルタ部の出力である属性名インデックスとを入力すると、前記属性名インデックスのうちの１つを取り出し、与えられた前記文書ファイルの内容中に、前記取り出した属性名を含むタグが存在するかどうか調べ、存在する場合は、その属性名のタグの値と前記入力した文書ファイル名とを利用者に表示し、存在しない場合には何も出力しないことを、前記入力した属性名インデックスのそれぞれに対して行う文書内容出力部、
前記合格文書ファイル名選別部の出力である文書ファイル名の集合を入力し、文書ファイル名格納辞書を参照して、前記入力した文書ファイル名の集合の要素を１つずつ取り出し、文書内容出力部に渡すことを、前記入力中の文書ファイル名のすべてに対して行うことを繰り返す文書内容出力制御部、
として機能させるプログラムを記録した機械読み取り可能な記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は文書ファイル検索装置に関し、特に自然言語による検索問い合わせを可能とした文書ファイル検索装置に関する。

【０００２】

【従来の技術】一般に情報検索において利用者の検索意図をより精密に表現させようとする場合には、日本語や英語のような自然言語によってそれを表現させる方法が有効である。データベースに対する検索を自然言語で行うシステムは既に存在し、自然言語インタフェースと呼ばれている（参考文献：ディベロッピングアナチュラルランゲージインタフェースツーコンプレックスデータ、ジー．ジー．ヘンドリックス他，“Developing a Natural Language Interface to Complex Data”，ACM Trans. on Database Systems, 1978.）。

【０００３】従来の自然言語インタフェースは、利用者の自然言語による検索問い合わせを解釈して、その問い合わせをデータベースの検索言語（SQL）の検索式に変換し、その検索式をデータベースシステムに送り、データベースシステムから戻された検索結果を利用者に提示するものである。

【０００４】しかし、従来の自然言語インタフェースは、既に商用化が始まって20年以上たったのにも関わらず、まだ実用のレベルに達していない。その理由の１つは、自然言語インタフェースシステムが利用者の自由な問い合わせを解釈することができず、そのシステムが許容する構文や語彙が明確に限定されているので、結局は利用者はどういう言い回しが使えるかを覚えなくてはならないためである。従って、自然言語インタフェースといっても複雑なコマンド体系と変わらない。つまり、従来の自然言語インタフェースは、利用者の自由な言い回しを受け付けることが出来ないと言うことが問題であった（参考文献：ディベロッピングアナチュラルランゲージインタフェースツーコンプレックスデータ、ジー．ジー．ヘンドリックス他，“Developing a Natural Langu

age Interface to Complex Data", ACM Trans. on Database Systems, 1978.)。とりわけ、簡単な質問については、利用者は完璧な自然言語でなく、簡潔な言い回しを使いたいのに、それが許されないということが問題であった。簡潔な言い回しの例としては、キーワード列による表現や非文法的な言い回し、あるいは自然言語文の一部、等がある。

【０００５】そこで本出願人は、先の特許出願において、簡潔な言い回しの自然言語による検索を実現する「自然言語解釈方法」を提案した（特開平５－６７１３６号公報）。これは、属性の属性名とその属性の属性値との組の集まりであるデータベースのテーブルを検索対象とし、自然言語による問い合わせ文中の各単語を属性名と属性値とその他とに分類し、属性名に分類された単語は応答属性名として保存し、属性値に分類された単語はその属性値とそれに対応する属性名とを組にして条件属性値組群として保存し、この保存された条件属性値組群中の属性値と属性名との組が全てテーブル中に存在する場合、前記応答属性名として保存された各属性名に対応する前記テーブル中の属性値を、問い合わせに対する回答として出力するものである。なお、これに類似する従来技術として、やはり本出願人によって先に出願された特開平５－２４２１４７号公報にかかる「自然言語解釈方法」がある。

【０００６】他方、最近におけるWorld Wide Web (WWW) の利用の拡大に伴い、WWW 上での検索技術の重要性が高まっている。WWW の利用者がWWW 上で情報検索をするときに使う典型的なツールはサーチエンジンである。この例としては、Altavista, Infoseek, Lycosなどが有名である。しかし、サーチエンジンでは、キーワードを組み合わせる検索の形式なので、利用者の検索の意図が直接的に反映させられないことが多い。例えば、ホテルに関する情報を探すときに、値段が150 ドルでペットを連れて行くのが可能で、しかもサンフランシスコ近辺にあるホテルのホームページを見つけたいときに、そのような検索意図をキーワードの並びのみで表現することは不可能である。仮に、(1) 式で表現するように

150 ドル、サンフランシスコ、ホテル、ペット可　　… (1)

とキーワードを並べたら、非常に大量のホームページのリストが出力される。

【０００７】自然言語インタフェースシステムを導入しWWW のホームページを検索できれば、上の例のような精密な検索条件を素直に表現し、適切なホームページのみが検索できることになる。しかしながら、WWW 上のホームページの検索に対し自然言語インタフェースを適用した例は見当たらない。

【０００８】なお、WWW のホームページに対する検索技術の他の例として、特開平10-40262号公報に記載された「情報検索装置」があるが、これは、感性表現データをキーワードにした検索を可能にすることで、明確な検索対象または検索条件を持たない利用者の感性に合った情報検索を目的としているため、上の例のような精密な検索条件による検索には向いていない。

【０００９】

【発明が解決しようとする課題】上述した特開平５－６７１３６号公報に記載された技術は、単純な方法で自然言語による問い合わせを解釈することができるので、自然言語インタフェースシステムを実用化する上で有効な手段となり得るが、未だ解決すべき課題が残されている。それは、自然言語による問い合わせ文中に或る属性名が存在する場合、それが必ず応答属性名として扱われ、問い合わせに対する回答中に含められるため、回答が冗長になる場合があることである。

【００１０】例えば、「属性名＝書名，その属性値＝人間失格、属性名＝著者，その属性値＝太宰治」を持つテーブルに対して、「書名＝人間失格の著者は？」という問い合わせを行った場合、まず、「書名」が属性名と判定されて応答属性名として保存され、次いで「人間失格」が属性値と判定されて属性値「人間失格」とその属性名である「書名」との組が条件属性値組群として保存され、次いで、「著者」が属性名と判定されて応答属性名として保存される。そして、条件属性値組群中の属性値「人間失格」と属性名「書名」との組を有する前記テーブルが検索され、そのテーブルから応答属性名「書名」と「著者」とに対応する属性値「人間失格」と「太宰治」とが検索されて出力される。つまり、「人間失格」をも出力している分、回答が冗長になっている。

【００１１】そこで本発明の目的は、自然言語による検索問い合わせに対する回答の冗長性を極力無くすことにある。

【００１２】また、本発明の他の目的は、WWW のホームページに対しても自然言語による検索問い合わせを可能にすることにある。

【００１３】

【課題を解決するための手段】（１）第１の発明

上述した特開平５－６７１３６号公報に記載された技術において、回答が冗長になっている理由は、自然言語による問い合わせ文中に属性名が存在する場合、それに対応する属性値を利用者が問い合わせ文中で記述しているにもかかわらず、一律に応答属性名として扱っているためである。そこで、本発明では、自然言語による問い合わせ文中で検索条件を指定するために或る属性名とそれに対応する属性値とを記述する場合、利用者はそれらを互いに隣接して記述する傾向にある点に着目し、同一の属性の属性名と属性値とが隣どうしに現れる場合にその属性名を応答属性名に含めないようにしている。より具体的には、属性の属性名とその属性の属性値との対を内蔵する文書ファイルを検索対象文書ファイルとし、検索

対象文書ファイルから、利用者が自然言語で指定した検索条件に適合する部分を検索する文書ファイル検索装置において、自然言語で表現した検索要求文を先頭から順に探査し、属性名を表現する自然言語表現に対してはその属性名を属性名インデックスとして出力し、属性値を表現する自然言語表現に対してはその属性値と属性名との対を属性値インデックスとして出力することを順次行うキーワード抽出部と、前記キーワード抽出部の出力を入力して先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する場合のみ、前記属性名インデックスを削除し、それ以外の部分はそのまま出力するキーワードフィルタ部と、検索対象文書ファイル中に、前記キーワードフィルタ部から出力された全ての属性値インデックスの属性名と属性値との対が内蔵されているか否かを調べ、内蔵されている場合、前記キーワードフィルタ部から出力された属性名インデックスの属性名に対応する属性値を検索対象文書ファイルから検索して出力する検索手段とを備えている。

【００１４】更に、検索要求文中のどの自然言語表現が属性名を表現し、またどの自然言語表現が属性値を表現しているかを正確に判定できるようにするために、検索対象文書ファイル中に存在する属性名について、属性名とその属性名を表現する自然言語表現との対を格納しておく属性名格納辞書と、検索対象文書ファイル中に存在する属性値について、属性値とその属性値に対応する属性名とその属性値を表現する自然言語表現との３つ組を格納しておく属性値格納辞書とを備え、前記キーワード抽出部は、自然言語で表現した検索要求文を先頭から順に探査し、属性名格納辞書を参照して、属性名を表現する自然言語表現が含まれていたら、その自然言語表現と対である属性名を属性名インデックスとして出力し、属性値格納辞書を参照して、属性値を表現する自然言語表現が含まれていたら、その自然言語表現と３つ組である属性値と属性名との対の集合を属性値インデックスとして出力する構成を有する。

【００１５】このように構成された本発明の文書ファイル検索装置にあっては、利用者が自然言語で表現した検索要求文を入力すると、まずキーワード抽出部が、検索要求文を先頭から順に探査し、属性名を表現する自然言語表現に対してはその属性名を属性名インデックスとして出力し、属性値を表現する自然言語表現に対してはその属性値と属性名との対を属性値インデックスとして出力し、次いでキーワードフィルタ部が、キーワード抽出部の出力を入力して先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する場合のみ、属性名インデックスを削除し、次いで、検索手段が、検索対象文書ファイル中に、キーワードフィルタ部から出力された全ての属性値インデックスの属性名と属性値との対が内蔵されているか否かを調べ、内蔵されている場合、キーワードフィルタ部から出力された属性名インデックスの属性名に対応する属性値を検索対象文書ファイルから検索して出力することにより、利用者への回答が冗長になるのを防いでいる。

【００１６】（２）第２の発明

WWW 上のホームページの検索に対し自然言語インタフェースを適用するのが困難であった理由は、WWW のホームページの中身が自然言語で書かれた文章や図から構成されるファイルであり、データベースのように、属性名と属性値の集合でないことである。つまり、従来の自然言語インタフェースシステムが対象とするデータベースは、属性名と属性値の集合だったため、従来の技術の項の例で出てくるホテルのデータベースがあるとすると、

（名前：Ｘホテル、値段：＄150 、ペット：可能、場所：サンフランシスコ）

（名前：Ｙホテル、値段：＄200 、ペット：不可、場所：ロサンゼルス）

（名前：Ｚホテル、値段：＄180 、ペット：不可、場所：シアトル）

のような形態で格納されており、利用者の問い合わせは、(2) 式で表現するように

「値段＝$150　かつ　ペット＝可能　かつ　場所＝サンフランシスコ」 … (2)

のように変換できた。これは、SQL 言語にそのまま変換される。しかし、WWW のホームページは、通常このような属性名と属性値の情報が入っていないので、SQL の式で表現できるような検索式に変換できない。つまり、従来のWWW のホームページ作成言語はHTMLである（参考文献：ワールドワイドウェブコンソーシアムのホームページ、URL http://www.w3.org)。HTMLでは、文書ファイル中に、その構成を表現するための属性名と属性値の組が内蔵されている。例えば、図６に示すのが、HTMLファイルの例である。ここで、＜と＞に囲まれたものが属性タグであり、単独で使われるもの（例：＜Ｂ＞）と、開始タグ（例：<TR>）と終了タグ（例：</TR >）の対で使われるものとがある。HTMLのタグの特徴は、それが文書ファイル中の外見の表現を定義するのに限定されていることである。例えば、表的な表現にするタグは＜TABLE ＞であり、改行を表すタグは＜Ｐ＞で表現される。このようなHTMLファイルをWWW ブラウザに読み込ませると、図７に示すような形態になってユーザに出力表示される。しかし、HTMLでは、文書中の意味を表現する為のタグを定義することは出来ない。

【００１７】そこで本発明では、WWW の文書ファイルに、その文書ファイル中の意味を表現する属性名と属性値との組を内蔵させる。具体的には、例えば、ファイル中に文書の内容を属性タグとその属性値の対の集合で表

現できるように拡張したXML (Extensible Markup Language)で文書を記述する（参考文献：ワールドワイドウェブコンソーシアムのホームページ、「エクステンシブル　マークアップ　ランゲージ 1.0」http://www.w3.org/TR/PR-xml-971208）。XML は、WWW の標準を決める機関であるワールドワイドウェブコンソーシアム（参考文献：ワールドワイドウェブコンソーシアムのホームページ、URL http://www.w3.org)によって1997年12月にその仕様が提案された。XML で記述された文書では、文書の内容を機械が可読になって内容による検索が可能になる。そこで、本発明ではそのことを利用してWWW のホームページに対して自然言語による検索問い合わせを実現する。

【００１８】具体的には、文書中に書かれた意味を表現する属性名のついたタグとその属性の属性値との対を内蔵する文書ファイルを検索対象文書ファイルとし、検索対象文書ファイルから、利用者が自然言語で指定した検索条件に適合する部分を検索する文書ファイル検索装置において、検索対象文書ファイル中に存在する属性名について、属性名とその属性名を表現する自然言語表現との対の集合を格納しておく属性名格納辞書と、検索対象文書ファイル中に存在する属性値について、属性値とその属性値に対応する属性名とその属性値を表現する自然言語表現との３つ組の集合を格納しておく属性値格納辞書と、自然言語で表現した検索要求文を先頭から順に探査し、属性名格納辞書を参照して、属性名を表現する自然言語表現が含まれていたら、その自然言語表現と対である属性名を属性名インデックスとして出力し、属性値格納辞書を参照して、属性値を表現する自然言語表現が含まれていたら、その自然言語表現と３つ組である属性値と属性名との対の集合を属性値インデックスとして出力するキーワード抽出部と、キーワード抽出部の出力を入力し、先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する場合のみ、前記属性名インデックスを削除し、それ以外の部分はそのまま出力するキーワードフィルタ部と、検索対象文書ファイル中に、前記キーワードフィルタ部から出力された全ての属性値インデックスの属性名と属性値との対に対応するタグが内蔵されているか否かを調べ、内蔵されている場合、前記キーワードフィルタ部から出力された属性名インデックスの属性名を持つタグの属性値を検索対象文書ファイルから検索して出力する検索手段とを備えている。

【００１９】このように構成された本発明の文書ファイル検索装置にあっては、利用者が自然言語で指定した検索要求文を入力すると、キーワード抽出部が、検索要求文を先頭から順に探査し、属性名を表現する自然言語表現が含まれていたら、その属性名を属性名インデックスとして出力し、属性値を表現する自然言語表現が含まれていたら、その属性値と属性名との対の集合を属性値インデックスとして出力し、次いで、キーワードフィルタ部が、キーワード抽出部の出力を入力し、先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する場合のみ、前記属性名インデックスを削除し、次いで、検索手段が、検索対象文書ファイル中に、キーワードフィルタ部から出力された全ての属性値インデックスの属性名と属性値との対に対応するタグが内蔵されているか否かを調べ、内蔵されている場合、キーワードフィルタ部から出力された属性名インデックスの属性名を持つタグの属性値を検索対象文書ファイルから検索して出力する。

【００２０】また、予め登録された多数の文書ファイルの内から利用者が自然言語で入力した検索条件を満たす文書ファイルのみを選別し、さらにその中の利用者が必要な部分を利用者に表示できるようにするために、文書中に書かれた意味を表現する属性名のついたタグとその属性の値との対を複数個内蔵する文書ファイルの集合から、利用者が自然言語で指定した検索条件を満足する文書ファイルを選択してその適合する部分を表示する文書ファイル検索装置において、検索対象となるすべての文書ファイルの名前と存在位置とを格納する文書ファイル名辞書と、検索対象となる文書ファイル中に存在する属性名について、属性名とその属性名を表現する自然言語表現との対の集合を格納しておく属性名格納辞書と、検索対象となる文書ファイル中に存在する属性値について、属性値とその属性値に対応する属性名とその属性値を表現する自然言語表現との３つ組の集合を格納しておく属性値格納辞書と、利用者が、自然言語で表現した検索要求文を入力すると、前記入力文を先頭から順に探査し、属性名格納辞書を参照して、属性名を表現する自然言語表現が含まれていたら、その自然言語表現と対である属性名を属性名インデックスとして出力し、属性値格納辞書を参照して、属性値を表現する自然言語表現が含まれていたら、その自然言語表現と３つ組である属性値と属性名との対の集合を属性値インデックスとして出力することを順次行うキーワード抽出部と、キーワード抽出部の出力を入力し、先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する場合のみ、前記属性名インデックスを削除し、それ以外の部分はそのまま出力するキーワードフィルタ部と、文書ファイルの内容と属性値インデックスとを入力すると、前記文書ファイルの内容中に、前記属性値インデックス中の属性名を含むタグが存在するかどうか調べ、存在する場合は、そのタグと対で存在する属性値を取り出し、その値が前記属性値インデックス中の属性値と等しいかどうか調べ、等しい場合は、合格の出力をし、そうでない場合は不合格の出力をする文書内容検査部と、文書ファイルの内容と１つ以上の属性値インデックスとを入力すると、前記属性値インデックスから１つずつ取り出し、前記文書ファイルの内容と前記取り出

した属性値インデックスを１つずつ文書内容検査部に渡していき、すべての属性値インデックスに対してその出力が合格のときは、合格を出力し、そうでないときは不合格を出力する統合文書内容検査部と、文書ファイル名辞書を参照して、１つずつ文書ファイルの内容を取り出し、前記文書の内容とキーワードフィルタ部の出力のうちの属性値インデックスの部分とを統合文書内容検査部に渡し、前記統合文書内容検査部の出力を受け取ることを前記１つずつ取り出した文書ファイルのすべてに対して行い、前記出力が合格の文書ファイルの名前のみを出力する合格文書ファイル名選別部と、文書ファイル名と前記文書ファイル名の内容とキーワードフィルタ部の出力である属性名インデックスとを入力すると、前記属性名インデックスのうちの１つを取り出し、与えられた前記文書ファイルの内容中に、前記取り出した属性名を含むタグが存在するかどうか調べ、存在する場合は、その属性名のタグの値と前記入力した文書ファイル名とを利用者に表示し、存在しない場合には何も出力しないことを、前記入力した属性名インデックスのそれぞれに対して行う文書内容出力部と、前記合格文書ファイル名選別部の出力である文書ファイル名の集合を入力し、文書ファイル名格納辞書を参照して、前記入力した文書ファイル名の集合の要素を１つずつ取り出し、文書内容出力部に渡すことを、前記入力中の文書ファイル名のすべてに対して行うことを繰り返す文書内容出力制御部とを備えている。

【００２１】このように構成された本発明の文書ファイル検索装置の作用を、その理解を容易にするために、例を使って説明する。まず、利用者が検索する対象となるWWW文書ファイルとして、図５（ａ），（ｂ）に示したものを使用する。図５の文書ファイル中には、文章テキストの他に、属性の属性名とその属性の属性値との対が含まれている。また、利用者の検索文の例として、次の文を使う。

検索入力文：「値段が150 ドルでペットを連れて行くのが可能で、しかも、サンフランシスコ近辺にあるホテルの情報を見つけたい」

【００２２】まず、第１段階では、入力文をキーワード列に変換する。キーワードの種類としては、２種類存在する。１つ目は、属性名を参照する自然言語表現であり、属性名インデックスと呼ぶ。２つ目は、属性値を参照する自然言語表現であり、属性値インデックスと呼ぶ。

【００２３】

生成されたキーワード列：
{
価格（「価格」の属性名インデックス）、
$150ドル（「価格」の属性値インデックス）、
ペット（「ペット可能性」の属性名インデックス）、
可能（「ペット可能性」の属性値インデックス）、
サンフランシスコ（「場所」の属性値インデックス）、
ホテル（「ホテル名」の属性名インデックス）
}　…(4)

【００２４】次に、属性名インデックスと属性値インデックスの並び順を参照して、冗長な部分の統合を行う。同一の属性に対する属性名インデックスと属性値インデックスとが隣りどうしに並んでいるときには、属性名インデックスの方を削除する。上のキーワード列例は、次のように圧縮される。

圧縮されたキーワード列：
{
$150ドル（「価格」の属性値インデックス）、
可能（「ペット可能性」の属性値インデックス）、
サンフランシスコ（「場所」の属性値インデックス）、
ホテル（「ホテル名」の属性名インデックス）
}　…(5)

【００２５】次に、抽出したキーワード列を解釈する。属性値インデックスは、それが参照する属性の値として、属性値インデックスが保持する値を取ること、という条件式と解釈する。例えば、

$150 ドル（「価格」の属性値インデックス）

は、

「「価格」属性の値　＝ $150 」

という解釈をする。

【００２６】複数の属性値インデックスが存在する場合は、それらの解釈を論理積したものが全体の条件式となる。上の例では、以下のようになる。

全体の条件式
{
「「価格」属性の値 = $150 」　かつ

「「ペット可能性」属性の値 = 可能」 かつ
「「場所」属性の値 = サンフランシスコ」
}
…(6)

【００２７】属性名インデックスは、それが参照する属性の値を出力せよ、という解釈になる。上の例では、以下のようになる。

検索部分の特定
{
ホテル（「ホテル名」の属性名インデックス）」
}
…(7)

【００２８】この意味は、「「ホテル名」属性の値を出力せよ」という解釈となる。複数の属性名インデックスがあるときは、それら複数の属性名インデックスを順次出力せよ、という意味になる。

【００２９】入力文全体の解釈は、属性値インデックスから生成される検索条件式を満足するWWW 上の文書ファイルを選択し、次に、それらの文書ファイル中から属性名インデックスの解釈で指定される属性名の値を抽出してそれを利用者に表示すれば良い。

【００３０】

【発明の実施の形態】図１を参照すると、本発明の実施の形態の文書ファイル検索装置１００は、文書ファイル名辞書１と、属性名格納辞書２と、属性値格納辞書３と、キーワード抽出部４と、キーワードフィルタ部５と、文書内容検査部６と、統合文書内容検査部７と、合格文書ファイル名選別部８と、文書内容出力部９と、文書内容出力制御部10とから構成され、キーボード等の入力装置１０１、ＣＲＴディスプレイ等の表示装置１０２およびインターネット１０３に接続されている。

【００３１】文書ファイル名辞書１には、検索対象となるすべての文書ファイルの名前とその物理的な位置とが格納されている。検索対象となる文書ファイルがHTMLやXMLで記述されている場合には、文書ファイルは、世界中のWWW サーバに分散していることも可能である。その場合、文書ファイルの位置は、「http://.....」というURL 記述になる。

【００３２】属性名格納辞書２には、検索対象となる文書ファイル中に存在する属性タグの属性名とその属性名を表現する自然言語表現との対が登録されている。ある属性名を参照する自然言語表現の中の最も基本的なものは、その属性名そのものである。例えば、「ホテル」という属性名を参照する自然言語表現としては、「ホテル」である。しかし、それ以外にも、「ホテル」を参照する表現がある。例えば、「宿泊場所」、「泊まるところ」などの表現がある。これらが、下記の表１でで示すような対になって登録される。

【００３３】

【表１】

属性名	自然言語表現
ホテル	ホテル
ホテル	宿泊場所
ホテル	泊まるところ

【００３４】属性値格納辞書３には、検索対象となる文書ファイル中に存在する属性値について、属性値とその属性値に対応する属性名とその属性値を表現する自然言語表現との３つ組が格納される。ある属性値を参照する自然言語表現としてもっとも基本的なものは、その属性値そのものである。例えば、「Ｘホテル」という属性値を参照する自然言語表現としては、「Ｘホテル」そのものがありこれ以外にはないかもしれない。しかし、別の例では、「ペット」属性の属性値を表わす自然言語表現としては、「可能」の他に「動物連れ込みOK」「ペット同伴OK」「犬猫可」のような表現も登録しておいてもよい。属性値格納辞書３には、下記の表２で示すように３つ組でデータが格納される。

【００３５】

【表２】

属性値	自然言語表現	対応する属性名
可能	可能	ペット
可能	動物連れ込みOK	ペット
可能	ペット同伴OK	ペット
可能	犬猫可	ペット
不可	動物不可	ペット

【００３６】キーワード抽出部４は、自然言語表現による入力条件検索文を入力装置101 を通じて利用者から受け取ると、属性名格納辞書２と属性値格納辞書３とを参照して、その中の自然言語表現として登録されている表現が入力条件検索文中にないかどうかを調べる。あった場合には、それが属性名の場合には、属性名のみを出力する。この出力のことを属性名インデックスと呼ぶ。他方、それが属性値の場合には、属性値と対応する属性名との対を出力する。この出力のことを属性値インデックスと呼ぶ。これらは、入力条件検索文の先頭から調べていき、マッチするものが見つかったら、その順番に出力していく。

【００３７】キーワードフィルタ部５は、キーワード抽出部４の出力をそのまま受け取り、先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する場合は、その属性名インデックスを削除し、それ以外の部分はそのまま素通しで出力する。

【００３８】文書内容検査部６は、統合文書内容検査部７から文書ファイルの内容である文字列と属性値インデックスとを入力として受け付ける。入力として受け取った文書ファイルの内容文字列中に、受け取った属性値インデックス中の属性名を含むタグが存在するかどうか調べ、存在する場合は、そのタグと対で存在する属性値を取り出し、その値がこの属性値インデックス中の属性値と等しいかどうか調べ等しい場合は、合格の出力をし、そうでない場合は不合格の出力をする。文書内容検査部６は、統合文書内容検査部７から呼び出されて動作する一種のサブルーチン的役割を果たしている。

【００３９】統合文書内容検査部７は、合格文書ファイル名選別部８から文書ファイルの内容である文字列と１つ以上の属性値インデックスとを入力として受け付ける。与えられた属性値インデックスは、１つ１つが「属性値インデックス中に記述された属性の値として、属性値インデックス中に記述された値をとらねばならない」という条件式を表現していると見做す。統合文書内容検査部７の役割は、与えられた文字列中から、属性値インデックスに記述された属性表現を見つけて、その条件が満足されているかを調べることである。入力として与えられた１つ以上の属性値インデックスのすべての条件を満足すれば、「合格」という値を出力し、そうでない場合は、「不合格」という値を出力する。実際に、文書ファイルの内容である文字列が１つの属性値インデックスの条件を満足するかどうかを判定するのは、文書内容検査部６が行う。統合文書内容検査部７は、複数の属性値インデックスがあった場合に、属性値インデックス１つずつを文書内容検査部６に順々に渡していく一種のループ制御を行っている。統合文書内容検査部７も、合格文書ファイル名選別部８から呼び出されるサブルーチン的役割である。

【００４０】合格文書ファイル名選別部８は、文書ファイル名辞書１を参照して、必要に応じてインターネット１０３を通じて世界中に分散しているWWW サーバをアクセスして１つずつ文書ファイルの内容を取り出し、この文書の内容とキーワードフィルタ部５の出力のうち属性値インデックスの部分とを統合文書内容検査部７に渡し、統合文書内容検査部７の出力を受け取る。ここで出力としては、「合格」または「不合格」の値が返される。この処理を文書ファイル名辞書１に登録されているすべてのファイルに対して行い、統合文書内容検査部７の出力が「合格」だったファイルに対してのみ、文書ファイル名を文書内容出力制御部10に出力する。

【００４１】文書内容出力部９は、文書ファイル名とこのファイルの内容とキーワードフィルタ部５の出力である１つ以上の属性名インデックスとを入力する。入力した属性名インデックスのうちの１つを取り出し、入力した文書ファイルの内容中に、この属性名インデックス中の属性名を含むタグが存在するかどうか調べ、存在する場合は、その属性名タグに対応する属性値タグの値と入力した文書ファイル名との対を表示装置102 を通じて利用者に表示し、存在しなかった場合には何も出力しないという処理を、入力したすべての属性名インデックスのそれぞれに対して行う。文書内容出力部９は、文書内容出力制御部10によってサブルーチン的に呼び出される役

割をしている。なお、属性値タグの値と文書ファイル名との対を出力する代わりに、属性値タグの値と文書ファイルの位置情報とを表示するようにしても良く、また、属性値タグの値と文書ファイル名とその位置情報とを表示するようにしても良い。

【００４２】文書内容出力制御部10は、合格文書ファイル名選別部８の出力である文書ファイル名の集合をそのまま自身の入力とし、文書ファイル名辞書１を参照して、入力した文書ファイル名の集合中の文書ファイルの内容を必要に応じてインターネット103 を通じてWWW サーバをアクセスして１つずつ取り出し、文書ファイル名およびキーワードフィルタ部５で生成された属性名インデックスとともに文書内容出力部９に渡すことを、入力中の文書ファイル名のすべてに対して行うことを繰り返すものである。つまり、入力として合格した文書ファイル名を３つ受け取った場合には、３回文書内容出力部９を呼び出すことになる。なお、合格文書ファイル名選別部８がインターネット103 を通じてWWWWサーバから取り込んだ文書ファイルの内容が磁気ディスク装置等に保存されている場合、文書内容出力制御部10はその内容を利用することで、インターネット103 へのアクセス回数を減らすことができる。

【００４３】図２および図３は文書ファイル検索装置100 の処理例を示すフローチャートである。以下、本実施の形態の動作について説明する。

【００４４】キーワード抽出部４は、入力装置101 を通じて利用者から自然言語表現による検索入力文を受け付けると（ステップＳ１）、属性名格納辞書２と属性値格納辞書３とを参照して、その中の自然言語表現として登録されている表現が検索入力文にないかどうかを、検索入力文の先頭から順に調べ、あった場合には、それが属性名のときは属性名のみを含む属性名インデックスを出力し、それが属性値のときは属性値と対応する属性名との対を含む属性値インデックスを出力する（ステップＳ２）。

【００４５】次にキーワードフィルタ部５は、キーワード抽出部４から出力されたインデックスの並びを検査し、同一の属性の属性名インデックスと属性値インデックスとが連続している箇所を検出し、その箇所の属性名インデックスを削除する（ステップＳ３）。

【００４６】次に合格文書ファイル名選別部８は、文書ファイル名辞書１中の１つの文書ファイル名に注目し、その文書ファイル名の文書の内容を取り出して、キーワードフィルタ部５から出力された全ての属性値インデックスとともに統合文書内容検査部７に渡し、合否を判定させる（ステップＳ４）。

【００４７】統合文書内容検査部７は、渡された文書内容を検査するために、まず渡された属性値インデックスの１つに注目し、この属性値インデックスと文書ファイルの内容とを文書内容検査部６に渡し、合否を判定させる（ステップＳ５）。

【００４８】文書内容検査部６は、渡された文書ファイルの内容中に、渡された属性値インデックスに含まれる属性名を持つ属性名タグが存在し、かつ、その存在した属性名タグと対になっている属性値タグの値が、渡された属性値インデックスに含まれる属性値と一致するかを検査し、一致する場合には合格を、そのような属性名タグが存在しないか或いは存在してもその属性値が一致しない場合には不合格を、統合文書内容検査部７に通知する（ステップＳ６）。

【００４９】統合文書内容検査部７は、文書内容検査部６から合格が通知された場合（ステップＳ７でＹＥＳ）、合格文書ファイル名選別部８から通知された全ての属性値インデックスについて検査し終えたか否かを調べ、未だ検査し終えていないときは（ステップＳ８でＮＯ）、残りの属性値インデックスの１つに注目を移し、その属性値インデックスと文書ファイルの内容とを文書内容検査部６に渡し、合否を判定させる（ステップＳ９）。そして、全ての属性値インデックスについて文書内容検査部７で合格の判定が出た場合（ステップＳ８でＹＥＳ）、合格文書ファイル名選別部８に合格を通知し、合格文書ファイル名選別部８は当該文書ファイルを合格文書ファイルとし（ステップＳ10）、ステップＳ11へと進む。他方、文書内容検査部６から不合格が通知された場合（ステップＳ７でＮＯ）、統合文書内容検査部７は合格文書ファイル名選別部８に不合格を通知し、合格文書ファイル名選別部８はステップＳ11へと進む。

【００５０】合格文書ファイル名選別部８は、１つの文書ファイルについての合否判定が終わると、文書ファイル名辞書１中に未処理の文書ファイルが残っている場合（ステップＳ11でＹＥＳ）、その内の１つの文書ファイル名に注目を移し（ステップＳ12）、先の文書ファイルと同様にして合否の判定を下す。

【００５１】文書ファイル名辞書１中の全ての文書ファイルに対する合否判定を終えると（ステップＳ11でＹＥＳ）、合格文書ファイル名選別部８は、少なくとも１つの合格ファイルがあったか否かを判定し（ステップＳ13）、１つもなければ、例えば入力された検索条件に合致する文書ファイルは１つもなかった旨を利用者に表示する等の処理を行って、処理を終了する。他方、１つでも合格ファイルが存在した場合、その全ての合格ファイルの文書ファイル名とキーワードフィルタ部５から出力された全ての属性名インデックスとを文書内容出力制御部10に通知して、文書内容出力制御を開始させる（ステップＳ14）。

【００５２】文書内容出力制御部10は、通知された１つの合格ファイル名に注目してその文書内容を取り出し、通知された全ての属性名インデックスとともに文書内容出力部９に渡し、当該文書の処理を開始させる（ステップＳ15）。

【００５３】文書内容出力部９は、通知された１つの属性名インデックスに注目し（ステップＳ16）、その属性名インデックスの属性名をもつ属性名タグが文書内にあるかを調べ（ステップＳ17）、あれば（ステップＳ18でＹＥＳ）、その属性名タグに対応する属性値タグの値と当該文書ファイル名とを表示装置１０２に表示する（ステップＳ19）。なければ（ステップＳ18でＮＯ）、ステップＳ19をスキップする。次に文書内容出力部９は、通知された属性名インデックスに未処理の属性名インデックスが残っているか否かを調べ（ステップＳ20）、残っていれば、その１つに注目を移し（ステップＳ21）、ステップＳ17に戻って上述した処理を繰り返す。

【００５４】文書内容出力部９が通知された全ての属性名インデックスについての処理を終えると（ステップＳ20でＮＯ）、文書内容出力制御部10は、合格文書ファイル名選別部８から通知された文書ファイルに未処理のものが残っているか否かを調べ（ステップＳ22）、残っている場合にはその１つに注目を移し、その文書ファイル名の文書内容を取り出して、合格文書ファイル名選別部８から通知された全ての属性名インデックスとともに文書内容出力部９に渡し、処理させる（ステップＳ23）。全ての合格ファイルについての処理が終わると（ステップＳ22でＹＥＳ）、処理終了となる。

【００５５】

【実施例】文書ファイル名辞書１に、図１に例示するように「ファイル１」，「ファイル２」，「ファイル３」の３つの文書ファイル名とそのＵＲＬとが登録されているとする。また、ファイル１の内容が図５（ａ）に示すものであり、ファイル２の内容が図５（ｂ）に示すものであるとする。これらのファイル１，２はXML で記述されており、文章テキストの他に属性と属性値が含まれている。つまり、ファイル１には、＜ホテル＞Ｘホテル＜／ホテル＞、＜場所＞サンフランシスコ＜／場所＞、＜値段＞＄１５０＜／値段＞、＜ペット＞可能＜／ペット＞といった、文書中に書かれた意味を表現する属性名のついたタグとその属性の値との対が含まれている。同様に、ファイル２にも、＜ホテル＞Ｚホテル＜／ホテル＞、＜場所＞シアトル＜／場所＞、＜値段＞＄１８０＜／値段＞、＜ペット＞不可＜／ペット＞といったタグが含まれている。

【００５６】また、属性名格納辞書２には図１に例示するような属性名とその自然言語表現との対が事前に格納されており、属性値格納辞書３には図１に例示するような属性値と自然言語表現と属性名との３つ組が事前に格納されているものとする。なお、属性値格納辞書３に全ての価格をその実際値で登録すると、登録数が増えてしまうので、変数を使用して登録するようにしても良い。つまり、XXX を任意の数値とする場合、以下の表３に示すように登録しておき、キーワード抽出部４は任意の数値の後ろに「ドル」があれば、自然言語表現XXX ドルが存在すると判断し、存在した実際値の頭に＄を付けたものを属性値とする。

【表３】

属性値	自然言語表現	属性名
$XXX	XXXドル	価格

【００５７】このような前提で、利用者が以下のような自然言語による検索入力文を入力した場合を例に、本実施例の動作を説明する。

検索入力文：「値段が150 ドルでペットを連れて行くのが可能で、しかも、サンフランシスコ近辺にあるホテルの情報を見つけたい」

【００５８】キーワード抽出部４は利用者からの検索入力文を受け付けると、属性名格納辞書２および属性値格納辞書３を参照して、検索入力文を以下のようにキーワード列に変換する。

【００５９】まず、検索入力文の先頭の自然言語表現「値段」が属性名格納辞書２に存在するので、それと対になって登録されている属性名「価格」を属性名インデックスとして出力する。次に、自然言語表現「150 ドル」が属性値格納辞書３に存在するので、それと３つ組で登録されている属性値「＄150 」と属性名「価格」との対を属性値インデックスとして出力する。次に、自然言語表現「ペット」が属性名格納辞書２に存在するので、それと対になって登録されている属性名「ペット」を属性名インデックスとして出力する。次に、自然言語表現「可能」が属性値格納辞書３に存在するので、それと３つ組で登録されている属性値「可能」と属性名「ペット」との対を属性値インデックスとして出力する。次に、自然言語表現「サンフランシスコ」が属性値格納辞書３に存在するので、それと３つ組で登録されている属性値「サンフランシスコ」と属性名「場所」との対を属性値インデックスとして出力する。次に、「ホテル」が属性名格納辞書２に存在するので、それと対になって登録されている属性名「ホテル」を属性名インデックスとして出力する。検索入力文中には、属性名格納辞書２および属性値格納辞書３に登録された自然言語表現とマッチする他の自然言語表現はない。従って、以下のようなキーワード列が上から順に出力される。

【００６０】属性名インデックス（属性名「価格」）
属性値インデックス（属性値「$150」、属性名「価格」）
属性名インデックス（属性名「ペット」）
属性値インデックス（属性値「可能」、属性名「ペット」）
属性値インデックス（属性値「「サンフランシスコ」、属性名「場所」）
属性名インデックス（属性名「ホテル」）

【００６１】次にキーワードフィルタ部５は、属性名インデックスと属性値インデックスとの並び順を参照して、冗長な部分の統合を行う。上のキーワード列の場合、属性名インデックス（属性名「価格」）と属性値インデックス（属性値「$150」、属性名「価格」）とは同じ属性名「価格」で隣どうしに並んでいるので、属性名インデックス（属性名「価格」）を削除する。また、属性名インデックス（属性名「ペット」）と属性値インデックス（属性値「可能」、属性名「ペット」）とは同じ属性名「ペット」で隣どうしに並んでいるので、属性名インデックス（属性名「ペット」）を削除する。他に削除すべき属性名インデックスは存在しないので、上記のキーワード列は最終的に以下のように圧縮される。

【００６２】(a) 属性値インデックス（属性値「$150」、属性名「価格」）

(b) 属性値インデックス（属性値「可能」、属性名「ペット」）

(c) 属性値インデックス（属性値「「サンフランシスコ」、属性名「場所」）

(d) 属性名インデックス（属性名「ホテル」）

【００６３】次に、合格文書ファイル名選別部８は、文書ファイル名辞書１中のファイル１の文書内容をそのＵＲＬを頼りにインターネット103 を通じて該当するサーバから取得し、その文書内容と上記の属性値インデックス(a) ～(c) とを統合文書内容検査部７に渡す。

【００６４】統合文書内容検査部７は、ファイル１の文書内容と、１つの属性値インデックス(a) とを文書内容検査部６に渡す。

【００６５】文書内容検査部６は、ファイル１の文書内容中に、属性値インデックス(a) 中の属性名「価格」のタグが存在するか否かを調べる。図５（ａ）のファイル１の場合、該当するタグ＜値段＞$150＜／値段＞があるので、その属性値「$150」が受け取った属性値インデックス(a) 中の属性値「$150」と一致するか否かを調べる。今の例では、一致するので、合格を統合文書内容検査部７に返却する。

【００６６】統合文書内容検査部７は、ファイル１の文書内容と、次の属性値インデックス(b) とを文書内容検査部６に渡す。

【００６７】文書内容検査部６は、ファイル１の文書内容中に、属性値インデックス(b) 中の属性名「ペット」のタグが存在するか否かを調べる。図５（ａ）のファイル１の場合、該当するタグ＜ペット＞可能＜／ペット＞があるので、その属性値「可能」が受け取った属性値インデックス(b) 中の属性値「可能」と一致するか否かを調べる。今の例では、一致するので、合格を統合文書内容検査部７に返却する。

【００６８】統合文書内容検査部７は、ファイル１の文書内容と、次の属性値インデックス(c) とを文書内容検査部６に渡す。

【００６９】文書内容検査部６は、ファイル１の文書内容中に、属性値インデックス(c) 中の属性名「場所」のタグが存在するか否かを調べる。図５（ａ）のファイル１の場合、該当するタグ＜場所＞サンフランシスコ＜／場所＞があるので、その属性値「サンフランシスコ」が受け取った属性値インデックス(c) 中の属性値「サンフランシスコ」と一致するか否かを調べる。今の例では、一致するので、合格を統合文書内容検査部７に返却する。

【００７０】統合文書内容検査部７は、ファイル１に関し全ての属性値インデックスで合格の結果が得られたので、合格文書ファイル名選別部８に合格を通知し、合格文書ファイル名選別部８はファイル１を合格ファイルとする。

【００７１】次に合格文書ファイル名選別部８は、文書ファイル名辞書１に格納されたファイル２の文書内容をそのＵＲＬを頼りにインターネット103 を通じて該当するサーバから取り込み、先のファイル１と同様に統合文書内容検査部７を使って合否を判定する。この場合、ペット属性、場所属性が満足しないので、ファイル２は不合格となる。同様に、残りのファイル３についても合否の判定が行われる。ここでは、ファイル３も不合格と判定され、合格ファイルはファイル１のみであったとする。

【００７２】次に合格文書ファイル名選別部８は、合格ファイル名としてファイル名１を、属性名インデックス(d) とともに文書内容出力制御部10に渡す。

【００７３】文書内容出力制御部10は、文書ファイル名辞書１からファイル名１のＵＲＬを取得し、それを頼りにインターネット103 上のサーバをアクセスしてファイル名１の文書内容を取得し、属性名インデックス(d) とともに文書内容出力部９に渡す。

【００７４】文書内容出力部９は、ファイル１の文書内容中に、属性名インデックス(d) の属性名「ホテル」を持つ属性タグが存在するか否かを調べる。図５（ａ）のファイル１の場合、該当するタグ＜ホテル＞Ｘホテル＜／ホテル＞が存在するので、その属性値「Ｘホテル」を取り出し、ファイル名１と共に表示装置102 に表示する。

【００７５】図４は本発明の文書ファイル検索装置のハードウェア構成例を示すブロック図である。この例の文書ファイル検索装置は、プロセッサ（ＣＰＵ）２００と、磁気ディスク装置等の補助記憶装置２０１と、そのインタフェース２０２と、ＲＡＭ等のメモリ２０３と、インターネット１０３との間のインタフェース２０４と、ＣＤ－ＲＯＭ，半導体メモリ等の機械読み取り可能な記録媒体２０５と、そのインタフェース２０６と、入力装置１０１と、そのインタフェース２０７と、表示装置１０２と、そのインタフェース２０８と、ＣＰＵ２００，メモリ２０３およびインタフェース２０２，２０

４，２０６～２０８間を接続するバス２０９とから構成されている。

【００７６】記録媒体２０５には、文書ファイル検索用プログラムが記録されており、このプログラムがインタフェース２０６を介してインストールされることにより、メモリ２０３または補助記憶装置２０１上に図１の文書ファイル名辞書１，属性名格納辞書２および属性値格納辞書３がロードされる。また、同プログラムはＣＰＵ２００の動作を制御することにより、ＣＰＵ２００を図１のキーワード抽出部４、キーワードフィルタ部５、文書内容検査部６、統合文書内容検査部７、合格文書ファイル名選別部８、文書内容出力部９、文書内容出力制御部10として機能させる。

【００７７】以上の実施の形態は本発明をWWW 上のホームページの検索に適用したが、特開平５－６７１３６号公報に記載する技術と同様にデータベースに対する検索に対しても適用可能である。

【００７８】

【発明の効果】以上説明したように本発明によれば以下のような効果が得られる。

【００７９】自然言語による検索問い合わせに対する回答の冗長性を極力無くすことができる。その理由は、キーワード抽出部で抽出されたインデックス列をその先頭から順に探査し、同一の属性の属性名インデックスと属性値インデックスとが隣どうしに存在する場合に属性名インデックスを削除するキーワードフィルタ部を備えているからである。

【００８０】WWW のホームページに対しても自然言語による検索問い合わせが可能になる。その理由は、XML のように文書ファイル中にその意味を表現する属性名と属性値のタグを内蔵させており、利用者が入力した自然言語による検索条件を解釈して適合する属性名および属性値をもつ文書ファイル中から、利用者の望む属性値を取り出すことができるからである。これによって、文法的に正しい自然言語表現、非文法的な表現、自然言語文の断片、キーワード列等、種々の形の入力を受け付けて統一的に解釈を行なう自然言語インタフェースによるWWW 文書検索システムを実現することが出来る。

【図面の簡単な説明】

【図１】本発明の実施の形態の文書ファイル検索装置のブロック図である。

【図２】本発明の実施の形態の文書ファイル検索装置処理例を示すフローチャートである。

【図３】本発明の実施の形態の文書ファイル検索装置処理例を示すフローチャートである。

【図４】本発明の文書ファイル検索装置のハードウェア構成例を示すブロック図である。

【図５】XML を使った文書ファイルの記述例を示す図である。

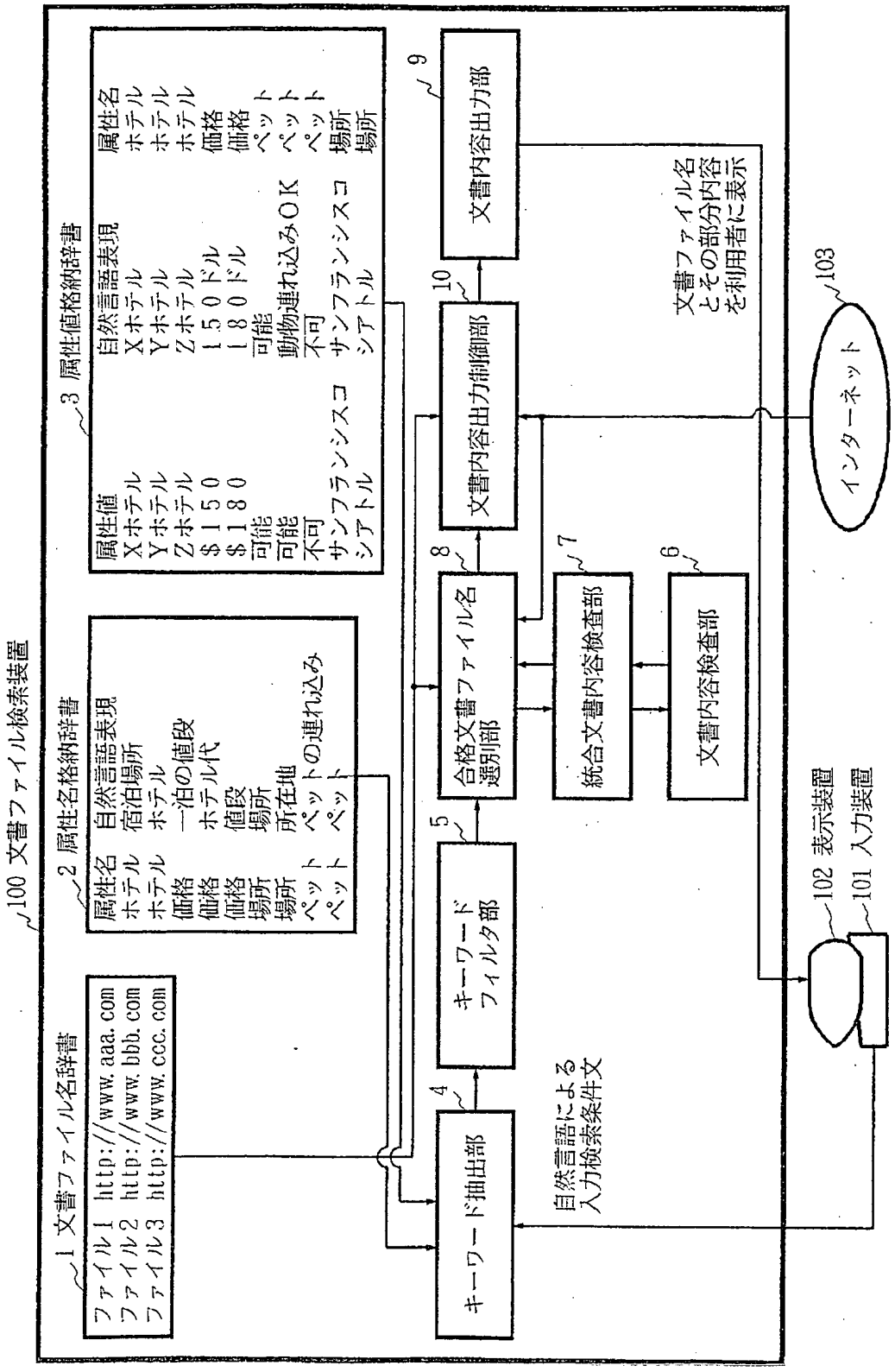
【図６】HTMLの記述例を示す図である。

【図７】図６のHTMLの記述例をブラウザで表示した例を示す図である。

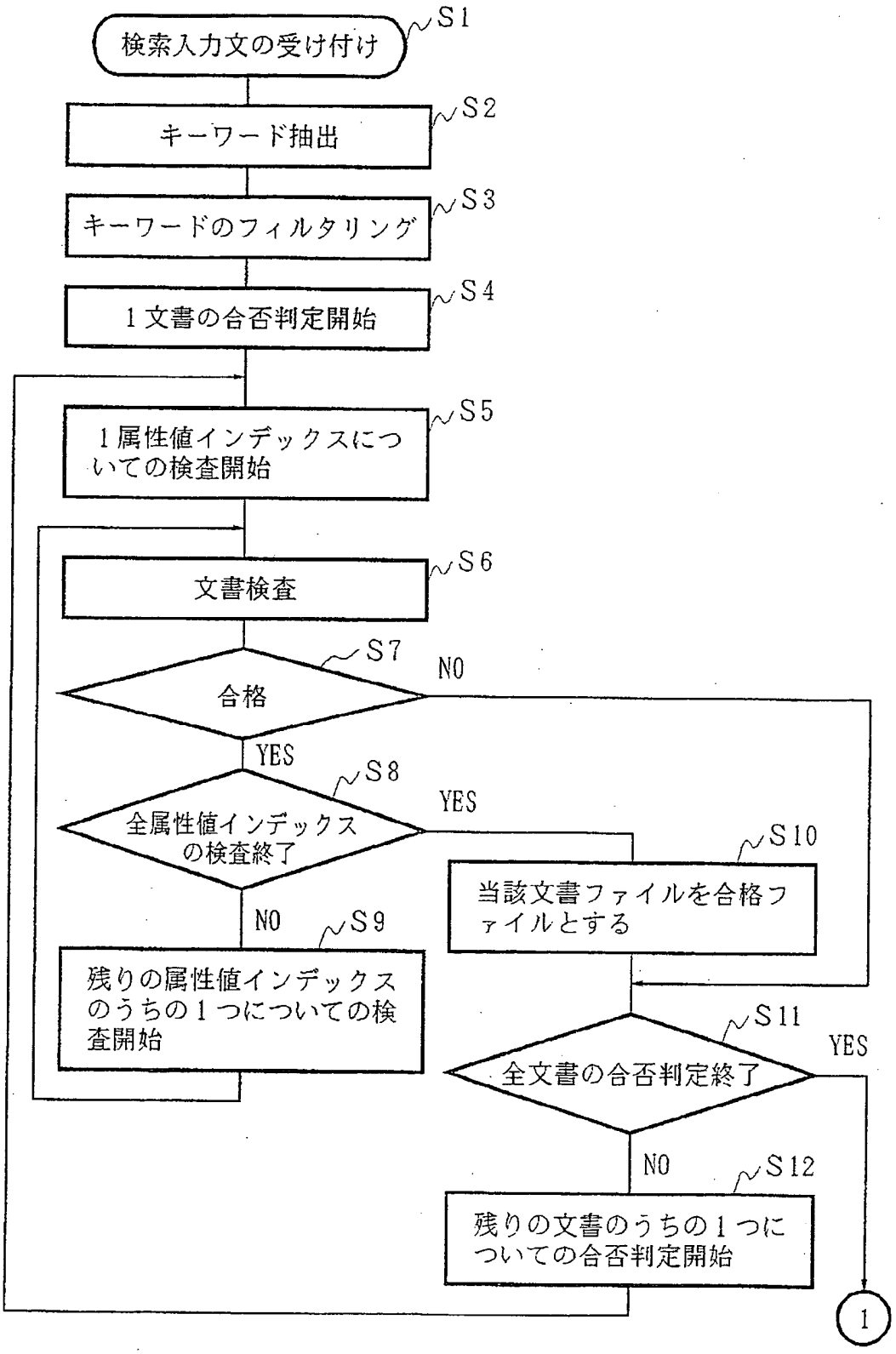
【符号の説明】

１は文書ファイル名辞書、２は属性名格納辞書、３は属性値格納辞書、４はキーワード抽出部、５はキーワードフィルタ部、６は文書内容検査部、７は統合文書内容検査部、８は合格文書ファイル名選別部、９は文書内容出力部、10は文書内容出力制御部、である。

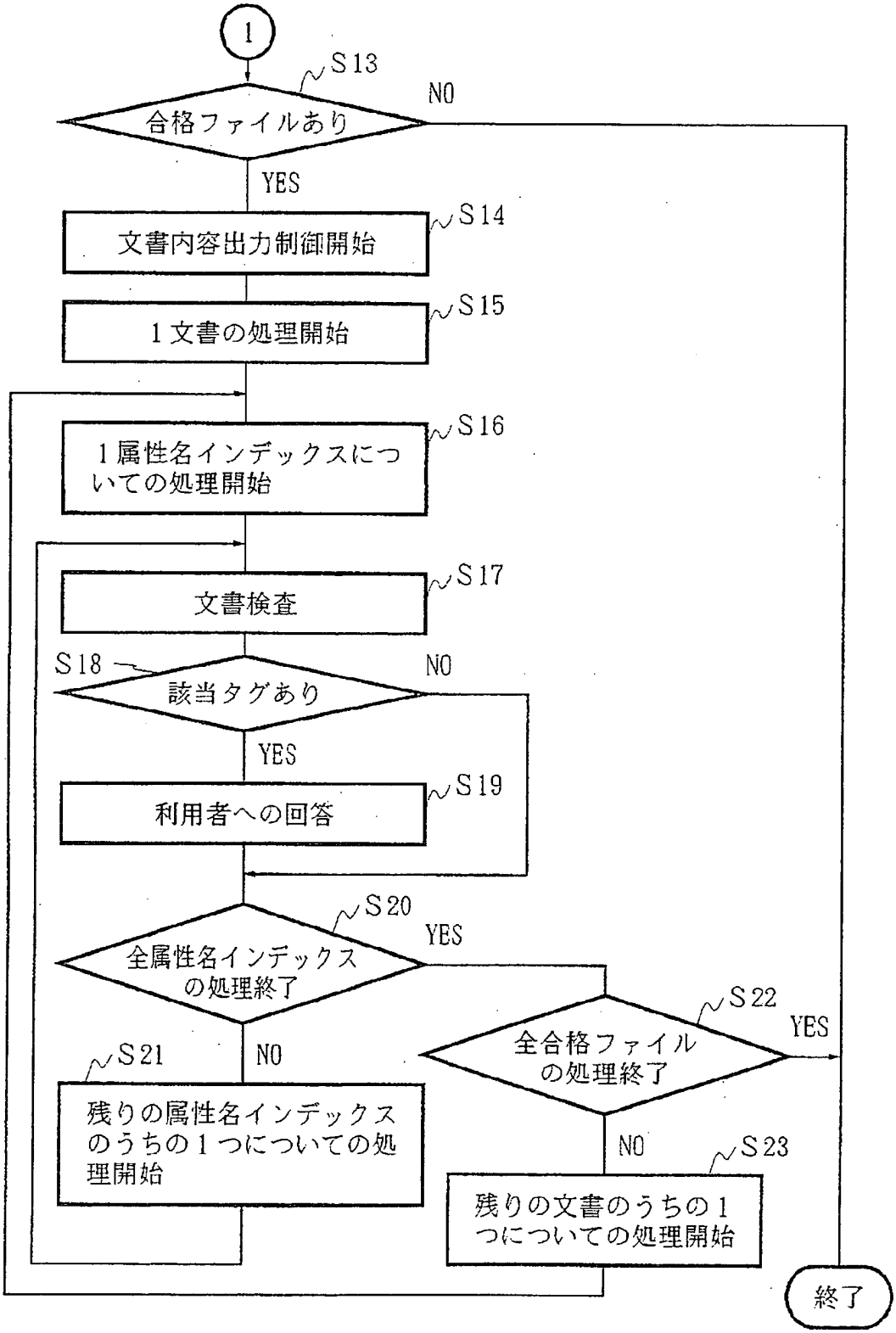
【図１】


100 文書ファイル検索装置
1 文書ファイル名辞書
ファイル1 http://www.aaa.com
ファイル2 http://www.bbb.com
ファイル3 http://www.ccc.com
2 属性名格納辞書
属性名 自然言語表現
ホテル 宿泊場所
ホテル ホテル
価格 一泊の値段
価格 ホテル代
価格 値段
場所 場所
場所 所在地
ペット ペットの連れ込み
ペット ペット
3 属性値格納辞書
属性値 自然言語表現 属性名
Xホテル Xホテル ホテル
Yホテル Yホテル ホテル
Zホテル Zホテル ホテル
$150 150ドル 価格
$180 180ドル 価格
可能 可能 ペット
可能 動物連れ込みOK ペット
不可 不可 ペット
サンフランシスコ サンフランシスコ 場所
シアトル シアトル 場所
4 キーワード抽出部
5 キーワードフィルタ部
6 文書内容検査部
7 統合文書内容検査部
8 合格文書ファイル名選別部
9 文書内容出力部
10 文書内容出力制御部
自然言語による入力検索条件文
文書ファイル名とその部分内容を利用者に表示
インターネット
102 表示装置
101 入力装置
103

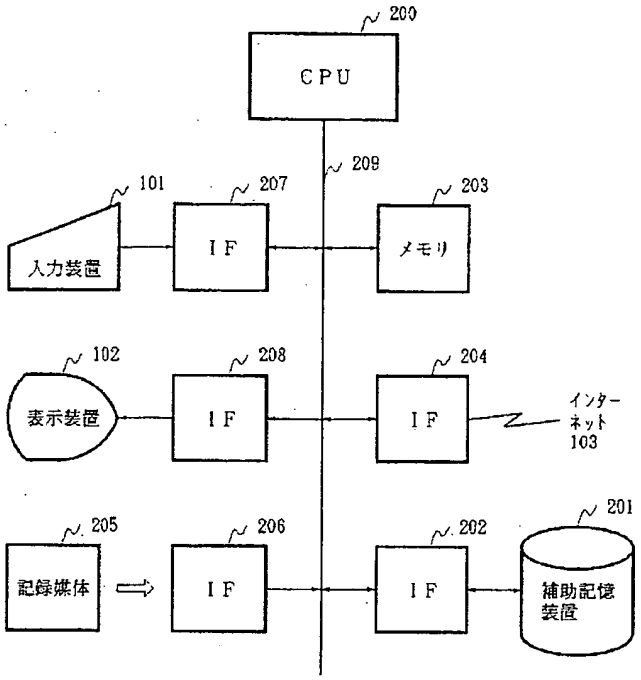
【図2】


検索入力文の受け付け
S1
キーワード抽出
S2
キーワードのフィルタリング
S3
1文書の合否判定開始
S4
1属性値インデックスについての検査開始
S5
文書検査
S6
合格
S7
NO
YES
全属性値インデックスの検査終了
S8
YES
NO
S9
残りの属性値インデックスのうちの1つについての検査開始
当該文書ファイルを合格ファイルとする
S10
全文書の合否判定終了
S11
YES
NO
S12
残りの文書のうちの1つについての合否判定開始
1

【図３】


1
S13
合格ファイルあり
NO
YES
S14
文書内容出力制御開始
S15
１文書の処理開始
S16
１属性名インデックスについての処理開始
S17
文書検査
S18
該当タグあり
NO
YES
S19
利用者への回答
S20
全属性名インデックスの処理終了
YES
S22
全合格ファイルの処理終了
YES
S21
NO
残りの属性名インデックスのうちの１つについての処理開始
NO
S23
残りの文書のうちの１つについての処理開始
終了

【図４】


200
CPU
209
101
207
203
入力装置
I F
メモリ
102
208
204
表示装置
I F
I F
インターネット
103
205
206
202
201
記録媒体
I F
I F
補助記憶装置

【図５】

(a)

```
<!--Xホテルのプロフィールのファイル-->  ⟸これはコメント行
<?xml version="1.0",encoding="shiftjis"?>
<!DOCTYPE 宿泊情報プロフィールSYSTEM"http://report/hotel.xml">
<プロフィール記録>
<ホテル>Xホテル</ホテル>
<場所>サンフランシスコ</場所>
<値段>$150</値段>
<ペット>可能</ペット>

<プロフィール>
ホテルXは、ビジネス街に位置し、ビジネス、観光いずれにも、・・・
</プロフィール>
・・・
</プロフィール記録>
```

(b)

```
<!--Zホテルのプロフィールのファイル-->  ⟸これはコメント行
<?xml version="1.0",encoding="shiftjis"?>
<!DOCTYPE 宿泊情報プロフィールSYSTEM"http://report/hotel.xml">
<プロフィール記録>
<ホテル>Zホテル</ホテル>
<場所>シアトル</場所>
<値段>$180</値段>
<ペット>不可</ペット>

<プロフィール>
ウォータフロントの中心部にあるZホテルは、シアトルで最もホットな場所として、・・・
</プロフィール>
・・・
</プロフィール記録>
```

【図６】

```
<html><title>1201</title>
<body BGCOLOR="#eeeeee" TBXT="#000000">
<TABLE BORDER="1" CELLSPACING="1" WIDTH="550">
<TR><TH>担当課題名</TH><TD COLSPAN="2">ＸＭＬの自然言語インタフェース方式の開発</TD></TR>
<TR><TH>担当名</TH><TD>ＮＥＣ　Ｃ＆Ｃメディア研究所</TD></TR>
<TR><TH>優先度</TH><TD>高い</TD></TR>
</TABLE>
<P>
<B>１．研究目的</B><P>
本研究は、先端的であり、・・・・
<P>
<B>２．具体的既往成果</B><P>
本研究部門では、すでに、自然言語インタフェースの研究を、・・・・
<P>
<B>３．全体計画</B><P>
１年目には、・・・・、２年目には、・・・・
</body></HTML>
```

担当課題名　　ＸＭＬの自然言語インタフェース方式の開発
担当名　　ＮＥＣ　Ｃ＆Ｃメディア研究所
優先度　　高い

１．研究目的

本研究は、先端的であり、・・・・

２．具体的既往成果

本研究部門では、すでに、自然言語インタフェースの研究を、・・・・

３．全体計画

１年目には、・・・・、２年目には、・・・・

【図７】

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
[Claim 1] In the text file retrieval equipment with which the text file which contains the pair of the attribute name and the attribute value of an attribute of an attribute is used as the text file for retrieval, and the part which suits the retrieval conditions which the user specified with natural language is searched from the text file for retrieval Investigate the retrieval demand sentence expressed with natural language sequentially from a head, and the attribute name is outputted as an attribute name index to the natural language expression expressing an attribute name. The keyword extraction section which performs outputting the pair of the attribute value and attribute name as an attribute value index to the natural language expression expressing attribute value one by one, Only when the output of said keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors The keyword filter section which deletes said attribute name index and outputs the other part as it is, It investigates whether the pair of the attribute name of all attribute value indexes and attribute value which were outputted from said keyword filter section into the text file for retrieval is built in. Text file retrieval equipment equipped with a retrieval means to search and output the attribute value corresponding to the attribute name of the attribute name index outputted from said keyword filter section from the text file for retrieval when built.
[Claim 2] The attribute name storing dictionary which stores the pair with the natural language expression expressing an attribute name and its attribute name about the attribute name which exists in the text file for retrieval, It has the attribute value storing dictionary which stores 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value, and its attribute value about the attribute value which exists in the text file for retrieval. Said keyword extraction section investigates the retrieval demand sentence expressed with natural language sequentially from a head, and refers to an attribute name storing dictionary. If the natural language expression expressing an attribute name is included, will output the attribute name which is the natural language expression and pair as an attribute name index, and an attribute value storing dictionary will be referred to. Text file retrieval equipment according to claim 1 characterized by having the configuration which outputs a pair of set with the natural language expression, the attribute value which is 3 groups, and an attribute name as an attribute value index if the natural language expression expressing attribute value is included.
[Claim 3] The text file which contains the pair of the tag which the attribute name expressing the semantics written into the document attached, and the attribute value of the attribute is used as the text file for retrieval. In the text file retrieval equipment with which the part which suits the retrieval conditions which the user specified with natural language from the text file for retrieval is searched The attribute name storing dictionary which stores a pair of set with the natural language expression expressing an attribute name and its attribute name about the attribute name which exists in the text file for retrieval, The attribute value storing dictionary which stores the set of 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value, and its attribute value about the attribute value which exists in the text file for retrieval, Investigate the retrieval demand sentence expressed with natural language sequentially from a head, and an attribute name storing dictionary is referred to. If the natural language expression expressing an attribute name is included, will output the attribute name which is the natural language expression and pair as an attribute name index, and an attribute value storing dictionary will be referred to. The keyword extraction section which will output the pair of the natural language expression, the attribute value which is 3 groups, and an attribute name as an attribute value index if the natural language expression expressing attribute value is included, Only when the output of the keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors The keyword filter section which deletes said attribute name index and outputs the other part as it is, It investigates whether the pair of the tag corresponding to the pair of the attribute name of all attribute value indexes and attribute value which were outputted

from said keyword filter section into the text file for retrieval is built in. Text file retrieval equipment equipped with a retrieval means to search and output the attribute value of the tag which has the attribute name of the attribute name index outputted from said keyword filter section when built from the text file for retrieval.

[Claim 4] From the set of the text file to build [two or more] in, the pair of the tag which the attribute name expressing the semantics written into the document attached, and the value of the attribute In the text file retrieval equipment which chooses the text file with which are satisfied of the retrieval conditions which the user specified with natural language, and displays the suiting part The text file term document which stores the identifier and existence location of all the text files used as the candidate for retrieval, The attribute name storing dictionary which stores a pair of set with the natural language expression expressing an attribute name and its attribute name about the attribute name which exists in the text file used as the candidate for retrieval, The attribute value storing dictionary which stores the set of 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value, and its attribute value about the attribute value which exists in the text file used as the candidate for retrieval, If a user inputs the retrieval demand sentence expressed with natural language, he will investigate said input statement sequentially from a head, and will refer to an attribute name storing dictionary. If the natural language expression expressing an attribute name is included, will output the attribute name which is the natural language expression and pair as an attribute name index, and an attribute value storing dictionary will be referred to. The keyword extraction section which performs one by one outputting a pair of set with the natural language expression, the attribute value which is 3 groups, and an attribute name as an attribute value index if the natural language expression expressing attribute value is included, Only when the output of the keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next door through The keyword filter section which deletes said attribute name index and outputs the other part as it is, When the contents and the attribute value index of a text file are inputted, in the contents of said text file Investigate whether a tag including the attribute name in said attribute value index exists, and when it exists Take out the attribute value which exists by the tag and pair, and whether the value is equal to the attribute value in said attribute value index investigation and when equal With the contents Banking Inspection Department of a document which outputs success, and does a rejected output when that is not right If the contents of the text file and one or more attribute value indexes are inputted Take out one at a time from said attribute value index, and the contents and said every one taken-out attribute value index of said text file are passed to the contents Banking Inspection Department of a document. To all attribute value indexes, when the output is success With the contents Banking Inspection Department of an integrated document which outputs success, and outputs a rejection when that is not right Every one contents of the text file are taken out with reference to said text file term document. The contents of said document, and the part of the attribute value index of the outputs of the keyword filter section to the contents Banking Inspection Department of an integrated document Delivery, The success text file name sorting section to which it performs receiving the output of said contents Banking Inspection Department of an integrated document to said all text files that it took out one [at a time], and said output outputs only the identifier of the text file of success, If the contents of a text file name and said text file name and the attribute name index which is the output of the keyword filter section are inputted When one of said attribute name indexes is taken out, and it investigates whether a tag including said taken-out attribute name exists and exists in the contents of said given text file with the contents output section of a document which said inputted attribute name index is alike, respectively, receives, and performs displaying the value and said inputted text file name of a tag of the attribute name on a user, and outputting nothing in not existing Input the set of the text file name which is the output of said success text file name sorting section, and a text file name storing dictionary is referred to. Text file retrieval equipment characterized by having the contents output-control section of a document which repeats performing taking out said every one inputted element of a set of a text file name, and passing the contents output section of a document to all the text file names under said input.

[Claim 5] The text file which contains the pair of the attribute name and the attribute value of an attribute of an attribute is used as the text file for retrieval. The computer which constitutes the text file retrieval equipment with which the part which suits the retrieval conditions which the user specified with natural language is searched from a text file for retrieval Investigate the retrieval demand sentence expressed with natural language sequentially from a head, and the attribute name is outputted as an attribute name index to the natural language expression expressing an attribute name. The keyword extraction section which performs outputting the pair of the attribute value and attribute name as an attribute value index to the natural language expression expressing attribute value one by one, Only when the output of said keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors The keyword filter section which deletes said attribute name index and outputs the other part as it is, It investigates whether the pair of the attribute name of all attribute value indexes and attribute value which were outputted from said keyword filter section into the text file for retrieval is built

in. The record medium which recorded the program which considers as a retrieval means to search and output the attribute value corresponding to the attribute name of the attribute name index outputted from said keyword filter section from the text file for retrieval when built, and is operated and in which machine reading is possible.

[Claim 6] The text file which contains the pair of the tag which the attribute name expressing the semantics written into the document attached, and the attribute value of the attribute is used as the text file for retrieval. The computer which constitutes the text file retrieval equipment with which the part which suits the retrieval conditions which the user specified with natural language is searched from a text file for retrieval The attribute name storing dictionary which stores a pair of set with the natural language expression expressing an attribute name and its attribute name about the attribute name which exists in the text file for retrieval, The attribute value storing dictionary which stores the set of 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value, and its attribute value about the attribute value which exists in the text file for retrieval, Investigate the retrieval demand sentence expressed with natural language sequentially from a head, and an attribute name storing dictionary is referred to. If the natural language expression expressing an attribute name is included, will output the attribute name which is the natural language expression and pair as an attribute name index, and an attribute value storing dictionary will be referred to. The keyword extraction section which will output the pair of the natural language expression, the attribute value which is 3 groups, and an attribute name as an attribute value index if the natural language expression expressing attribute value is included, Only when the output of the keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors The keyword filter section which deletes said attribute name index and outputs the other part as it is, It investigates whether the pair of the tag corresponding to the pair of the attribute name of all attribute value indexes and attribute value which were outputted from said keyword filter section into the text file for retrieval is built in. The record medium which recorded the program which considers as a retrieval means to search and output the attribute value of the tag which has the attribute name of the attribute name index outputted from said keyword filter section when built from the text file for retrieval, and is operated and in which machine reading is possible.

[Claim 7] From the set of the text file to build [two or more] in, the pair of the tag which the attribute name expressing the semantics written into the document attached, and the value of the attribute The computer which constitutes the text file retrieval equipment which chooses the text file with which are satisfied of the retrieval conditions which the user specified with natural language, and displays the suiting part The text file term document which stores the identifier and existence location of all the text files used as the candidate for retrieval, The attribute name storing dictionary which stores a pair of set with the natural language expression expressing an attribute name and its attribute name about the attribute name which exists in the text file used as the candidate for retrieval, The attribute value storing dictionary which stores the set of 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value, and its attribute value about the attribute value which exists in the text file used as the candidate for retrieval, If a user inputs the retrieval demand sentence expressed with natural language, he will investigate said input statement sequentially from a head, and will refer to an attribute name storing dictionary. If the natural language expression expressing an attribute name is included, will output the attribute name which is the natural language expression and pair as an attribute name index, and an attribute value storing dictionary will be referred to. The keyword extraction section which performs one by one outputting a pair of set with the natural language expression, the attribute value which is 3 groups, and an attribute name as an attribute value index if the natural language expression expressing attribute value is included, Only when the output of the keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next door through Delete said attribute name index, and if the keyword filter section, and the contents and the attribute value index of a text file outputted as it is are inputted, the other part When it investigates whether a tag including the attribute name in said attribute value index exists and exists in the contents of said text file Take out the attribute value which exists by the tag and pair, and whether the value is equal to the attribute value in said attribute value index investigation and when equal If success is outputted, and the contents Banking Inspection Department of a document, and the contents of the text file and one or more attribute value indexes which carry out a rejected output are inputted when that is not right Take out one at a time from said attribute value index, and the contents and said every one taken-out attribute value index of said text file are passed to the contents Banking Inspection Department of a document. To all attribute value indexes, when the output is success The contents Banking Inspection Department of an integrated document which outputs success, and outputs a rejection when that is not right, and said text file term document are referred to. Every one contents of the text file are taken out. The contents of said document, and the part of the attribute value index of the outputs of the keyword filter section to the contents Banking Inspection Department of an integrated document Delivery, It performs receiving the output of said contents Banking Inspection Department of an integrated document to said all text files that it took out

one [at a time]. If the attribute name index said whose output is an output of the success text file name sorting section which outputs only the identifier of the text file of success, and the contents of a text file name and said text file name and the keyword filter section is inputted When one of said attribute name indexes is taken out, and it investigates whether a tag including said taken-out attribute name exists and exists in the contents of said given text file The value and said inputted text file name of a tag of the attribute name are displayed on a user. the contents output section of a document which said inputted attribute name index is alike, respectively, receives, and performs outputting nothing in not existing -- Input the set of the text file name which is the output of said success text file name sorting section, and a text file name storing dictionary is referred to. The record medium which recorded the program which makes it the contents output-control section of a document to repeat to perform taking out said every one inputted element of a set of a text file name, and passing the contents output section of a document to all the text file names under said input, and is operated and in which machine reading is possible.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] Especially this invention relates to the text file retrieval equipment which made the retrieval inquiry by natural language possible about text file retrieval equipment.
[0002]
[Description of the Prior Art] When it is going to make a precision express a retrieval intention of a user more in information retrieval generally, the method of making it express with natural language like Japanese or English is effective. The system which performs retrieval to a database with natural language already exists, and is called the natural language interface (bibliography: DIBEROPPINGUANACHURARURANGEJI interface two complex data and G . G . "Developing a Natural Language Interface to Complex Data" besides Hendrix, ACM Trans.on Database Systems, 1978.).
[0003] The conventional natural language interface interprets the retrieval inquiry by a user's natural language, changes the inquiry into the retrieval type of the query language (SQL) of a database, and shows a user the retrieval result to which the retrieval type was returned by database system from delivery and database system.
[0004] However, although commercialization already started and the conventional natural language interface left for 20 years or more, it has not reached the level of practical use yet. Since the functor in which a natural language interface system cannot interpret a free inquiry of a user, but the system permits one of the reason of the, and a vocabulary are limited clearly, a user is for having to memorize what kind of expression can be used after all. Therefore, a natural language interface is not different from a complicated command system. That is, it was a problem to say that the conventional natural language interface cannot receive a user's free expression (bibliography: DIBEROPPINGUANACHURARURANGEJI interface two complex data and G . G . "Developing a Natural Language Interface to Complex Data" besides Hendrix, ACM Trans.on Database Systems, 1978.). It was especially a problem that it is not allowed although he wants to use not the natural language with a perfect user about an easy question but a brief expression. As an example of a brief expression, there are a part of expression by the keyword train, nongrammatical expression or natural language sentence, and **.
[0005] Then, these people proposed the "natural language interpretation approach" of realizing retrieval by the natural language of a brief expression, in previous patent application (JP,5-67136,A). This makes applicable to retrieval the table of the database which is the assembly of the group of the attribute name and the attribute value of an attribute of an attribute. Each word in the inquiry sentence by natural language is classified into an attribute name, attribute value, and others. Save the word classified into the attribute name as a response attribute name, and the word classified into attribute value makes a group the attribute value and the attribute name corresponding to it, and it saves it as condition attribute value ****. When all the groups of the attribute value in this saved condition attribute value **** and an attribute name exist in a table, the attribute value in said table corresponding to each attribute name saved as said response attribute name is outputted as a reply to an inquiry. In addition, there is the "natural language interpretation approach" which starts JP,5-242147,A for which these people applied previously too as a conventional technique similar to this.
[0006] On the other hand, it follows on expansion of use of World Wide Web (WWW) in recently, and is WWW. The importance of the upper retrieval technique is increasing. WWW A user is WWW. The typical tool used when carrying out information retrieval in a top is a search engine. As this example, Altavista, Infoseek, Lycos, etc. are famous. However, in a search engine, since it is the format of the retrieval which combines a keyword, the intention of retrieval of a user is not reflected directly in many cases. For example, a price is 150 when looking for the information about a hotel. It is possible to take a pet in a dollar, and it is impossible to express such a retrieval intention only along a keyword

to find the homepage of the hotel which is moreover near San Francisco. Temporarily, it is (1). It expresses by the formula. 150 A dollar, San Francisco, a hotel, and a pet are possible. -- If (1) and a keyword are put in order, the list of very a lot of homepages will be outputted.

[0007] A natural language interface system is introduced and it is WWW. If a homepage can be searched, precise retrieval conditions like the upper example will be expressed obediently, and only a suitable homepage can be searched. However, WWW The example which applied the natural language interface to retrieval of the upper homepage is not found.

[0008] in addition, WWW although there be "information retrieval equipment" indicated by JP,10-40262,A as other examples of the retrieval technique over a homepage, since this aim at the information retrieval suitable for the sensibility of the user who do not have the retrieval conditions for [clear] retrieval by enable retrieval which made sensibility expression data the keyword, it be fit for retrieval by precise retrieval conditions like the upper example.

[0009]

[Problem(s) to be Solved by the Invention] Although the technique indicated by JP,5-67136,A mentioned above can serve as an effective means when putting a natural language interface system in practical use since it can interpret the inquiry by natural language by the simple approach, the technical problem which should still be solved is left behind. Since it is surely treated as a response attribute name and is included during the reply to an inquiry when a certain attribute name exists in the inquiry sentence by natural language, it is that a reply may become redundancy.

[0010] For example, a table with "an attribute name = title, its attribute value = No Longer Human, an attribute name = author, and its attribute value = Dazaiosamu" is received. When the inquiry "the author of title = No Longer Human is ?" is performed, first Subsequently, "No Longer Human" is judged to be attribute value, a group with the "title" which is attribute value "No Longer Human" and its attribute name is saved as condition attribute value ****, a "title" is judged to be an attribute name, and it is saved as a response attribute name, and he is saved [an "author" is judged to be an attribute name and] as a response attribute name. And said table which has the attribute value in condition attribute value **** "No Longer Human" and a group with an attribute name "a title" is searched, and the attribute value "No Longer Human" and "Dazaiosamu" corresponding to a response attribute name "a title" and an "author" are searched and outputted from the table. That is, the part and reply which are outputting "No Longer Human" are redundancy.

[0011] Then, the purpose of this invention is to abolish the redundancy of the reply to the retrieval inquiry by natural language as much as possible.

[0012] Moreover, other purposes of this invention are WWW. It is in making the retrieval inquiry by natural language possible also to a homepage.

[0013]

[Means for Solving the Problem] (1) In the technique indicated by JP,5-67136,A in which the 1st carried out invention ****, although the user asked the attribute value corresponding to it and the reason the reply is redundancy has described it in a sentence when an attribute name exists in the inquiry sentence by natural language, it is because it is treating as a response attribute name uniformly. Then, when the attribute name and attribute value of the same attribute appear in next doors, he is trying not to include the attribute name in a response attribute name in this invention paying attention to the point in the inclination which a user adjoins them mutually and is described when describing a certain attribute name and the attribute value corresponding to it, in order to specify retrieval conditions in the inquiry sentence by natural language. The text file which contains the pair of the attribute name and the attribute value of an attribute of an attribute is more specifically used as the text file for retrieval. In the text file retrieval equipment with which the part which suits the retrieval conditions which the user specified with natural language from the text file for retrieval is searched Investigate the retrieval demand sentence expressed with natural language sequentially from a head, and the attribute name is outputted as an attribute name index to the natural language expression expressing an attribute name. The keyword extraction section which performs outputting the pair of the attribute value and attribute name as an attribute value index to the natural language expression expressing attribute value one by one, Only when the output of said keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors The keyword filter section which deletes said attribute name index and outputs the other part as it is, It investigates whether the pair of the attribute name of all attribute value indexes and attribute value which were outputted from said keyword filter section into the text file for retrieval is built in. When built, it has a retrieval means to search and output the attribute value corresponding to the attribute name of the attribute name index outputted from said keyword filter section from the text file for retrieval.

[0014] Furthermore, in order to enable it to judge correctly which natural language expression which natural language expression in a retrieval demand sentence expresses an attribute name, and is expressing attribute value The attribute name storing dictionary which stores the pair with the natural language expression expressing an attribute name and its

attribute name about the attribute name which exists in the text file for retrieval, It has the attribute value storing dictionary which stores 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value, and its attribute value about the attribute value which exists in the text file for retrieval. Said keyword extraction section investigates the retrieval demand sentence expressed with natural language sequentially from a head, and refers to an attribute name storing dictionary. If the natural language expression expressing an attribute name is included, will output the attribute name which is the natural language expression and pair as an attribute name index, and an attribute value storing dictionary will be referred to. If the natural language expression expressing attribute value is included, it has the configuration which outputs a pair of set with the natural language expression, the attribute value which is 3 groups, and an attribute name as an attribute value index.

[0015] Thus, if it is in the text file retrieval equipment of constituted this invention When a user inputs the retrieval demand sentence expressed with natural language, first the keyword extraction section Investigate a retrieval demand sentence sequentially from a head, and the attribute name is outputted as an attribute name index to the natural language expression expressing an attribute name. To the natural language expression expressing attribute value, the pair of the attribute value and attribute name is outputted as an attribute value index. Only when the keyword filter section inputs the output of the keyword extraction section, and investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors, an attribute name index is deleted. Subsequently, subsequently It investigates whether the pair of the attribute name of all the attribute value indexes with which the retrieval means was outputted from the keyword filter section into the text file for retrieval, and attribute value is built in. When built, it has prevented the reply to a user becoming redundancy by searching and outputting the attribute value corresponding to the attribute name of the attribute name index outputted from the keyword filter section from the text file for retrieval.

[0016] (2) 2nd invention WWW For applying a natural language interface to retrieval of the upper homepage, the difficult reason is WWW. The contents of the homepage are the files which consist of the texts and drawings which were drawn with natural language, and it is not being the set of an attribute name and attribute value like a database. That is, the target database [interface system / which is the former / natural language] is supposing there is a database of the hotel which comes out in the example of the term of a Prior art since it was the set of an attribute name and attribute value (identifier: X hotel, price:$150, pet:possibility of, location:San Francisco),
(Identifier: Y hotel, price:$200, pet:improper, location:Los Angeles)
(Identifier: Z hotel, price:$180, pet:improper, location:Seattle)
** -- it stores with a gestalt [like] -- having -- **** -- an inquiry of a user -- (2) it expresses by the formula -- as -- "-- price =$150 and pet = -- possible and location = San Francisco"
-- (2) It was [like] convertible. This is SQL. It is changed into language as it is. However, WWW Since the information on such [usually] an attribute name and attribute value is not contained, a homepage is SQL. It is not convertible for the retrieval type which can be expressed by the formula. That is, the conventional WWW Homepage creation language is HTML (bibliography: the homepage of a World-Wide-Web consortium, URL http://www.w3.org). In HTML, the group of the attribute name for expressing the configuration and attribute value is built in in the text file. For example, it is the example of an HTML file which is shown in drawing 6 . here -- < -- what it is the attribute tag which was surrounded by > and used independently (example : <B>) Initiation tag (example : <TR>) Termination tag (example : </TR>) There are some which are used by the pair. The description of a HTML tag is limited to it defining the expression of the appearance in a text file. For example, the tag made a table-expression is <TABLE> and the tag showing line feed is expressed by <P>. It is WWW about such an HTML file. If it is made to read into a browser, it will become a gestalt as shown in drawing 7 , and will be indicated by the output at a user. However, HTML cannot define the tag for expressing the semantics in a document.

[0017] So, at this invention, it is WWW. The group of the attribute name and attribute value expressing the semantics in the text file is made to build in a text file. XML (Extensible Markup Language) specifically extended so that the contents of the document could be expressed by a pair of set of an attribute tag and its attribute value in a file describes a document (bibliography: the homepage of a World-Wide-Web consortium, "EKUSUTENSHIBURU markup language 1.0"http://www.w3.org/TR/PR-xml -971208). . XML WWW The specification was proposed by the World-Wide-Web consortium (bibliography: the homepage of a World-Wide-Web consortium, URL http://www.w3.org) which is the engine which decides a criterion in December, 1997. XML By the described document, a machine becomes legible about the contents of the document and a content addressed retrieval becomes possible. Then, that is used in this invention and it is WWW. The retrieval inquiry by natural language is realized to a homepage.

[0018] The text file which contains the pair of the tag which the attribute name which specifically expresses the semantics written into the document attached, and the attribute value of the attribute is used as the text file for retrieval.

In the text file retrieval equipment with which the part which suits the retrieval conditions which the user specified with natural language from the text file for retrieval is searched The attribute name storing dictionary which stores a pair of set with the natural language expression expressing an attribute name and its attribute name about the attribute name which exists in the text file for retrieval, The attribute value storing dictionary which stores the set of 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value, and its attribute value about the attribute value which exists in the text file for retrieval, Investigate the retrieval demand sentence expressed with natural language sequentially from a head, and an attribute name storing dictionary is referred to. If the natural language expression expressing an attribute name is included, will output the attribute name which is the natural language expression and pair as an attribute name index, and an attribute value storing dictionary will be referred to. The keyword extraction section which will output a pair of set with the natural language expression, the attribute value which is 3 groups, and an attribute name as an attribute value index if the natural language expression expressing attribute value is included, Only when the output of the keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors The keyword filter section which deletes said attribute name index and outputs the other part as it is, It investigates whether the tag corresponding to the pair of the attribute name of all attribute value indexes and attribute value which were outputted from said keyword filter section into the text file for retrieval is built in. When built, it has a retrieval means to search and output the attribute value of a tag with the attribute name of the attribute name index outputted from said keyword filter section from the text file for retrieval.

[0019] Thus, if it is in the text file retrieval equipment of constituted this invention When a user inputs the retrieval demand sentence specified with natural language, the keyword extraction section If the natural language expression which investigates a retrieval demand sentence sequentially from a head, and expresses an attribute name is included The attribute name is outputted as an attribute name index, and if the natural language expression expressing attribute value is included, a pair of set with the attribute value and attribute name will be outputted as an attribute value index. Subsequently Only when the keyword filter section inputs the output of the keyword extraction section, and investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors, said attribute name index is deleted. Subsequently It investigates whether the tag corresponding to the pair of the attribute name of all the attribute value indexes with which the retrieval means was outputted from the keyword filter section into the text file for retrieval, and attribute value is built in. When built, the attribute value of a tag with the attribute name of the attribute name index outputted from the keyword filter section is searched and outputted from the text file for retrieval.

[0020] Moreover, in order to sort out only the text file which fulfills the retrieval conditions which the user inputted with natural language from from among the text files of a large number registered beforehand and to enable it to display the part which the user in it still needs on a user From the set of the text file to build [two or more] in, the pair of the tag which the attribute name expressing the semantics written into the document attached, and the value of the attribute In the text file retrieval equipment which chooses the text file with which are satisfied of the retrieval conditions which the user specified with natural language, and displays the suiting part The text file term document which stores the identifier and existence location of all the text files used as the candidate for retrieval, The attribute name storing dictionary which stores a pair of set with the natural language expression expressing an attribute name and its attribute name about the attribute name which exists in the text file used as the candidate for retrieval, The attribute value storing dictionary which stores the set of 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value, and its attribute value about the attribute value which exists in the text file used as the candidate for retrieval, If a user inputs the retrieval demand sentence expressed with natural language, he will investigate said input statement sequentially from a head, and will refer to an attribute name storing dictionary. If the natural language expression expressing an attribute name is included, will output the attribute name which is the natural language expression and pair as an attribute name index, and an attribute value storing dictionary will be referred to. The keyword extraction section which performs one by one outputting a pair of set with the natural language expression, the attribute value which is 3 groups, and an attribute name as an attribute value index if the natural language expression expressing attribute value is included, Only when the output of the keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors The keyword filter section which deletes said attribute name index and outputs the other part as it is, When the contents and the attribute value index of a text file are inputted, in the contents of said text file Investigate whether a tag including the attribute name in said attribute value index exists, and when it exists Take out the attribute value which exists by the tag and pair, and whether the value is equal to the attribute value in said attribute value index investigation and when equal With the contents Banking Inspection Department of a document which outputs success, and does a rejected output

when that is not right If the contents of the text file and one or more attribute value indexes are inputted Take out one at a time from said attribute value index, and the contents and said every one taken-out attribute value index of said text file are passed to the contents Banking Inspection Department of a document. To all attribute value indexes, when the output is success With the contents Banking Inspection Department of an integrated document which outputs success, and outputs a rejection when that is not right Every one contents of the text file are taken out with reference to a text file term document. The contents of said document, and the part of the attribute value index of the outputs of the keyword filter section to the contents Banking Inspection Department of an integrated document Delivery, The success text file name sorting section to which it performs receiving the output of said contents Banking Inspection Department of an integrated document to said all text files that it took out one [at a time], and said output outputs only the identifier of the text file of success, If the contents of a text file name and said text file name and the attribute name index which is the output of the keyword filter section are inputted When one of said attribute name indexes is taken out, and it investigates whether a tag including said taken-out attribute name exists and exists in the contents of said given text file with the contents output section of a document which said inputted attribute name index is alike, respectively, receives, and performs displaying the value and said inputted text file name of a tag of the attribute name on a user, and outputting nothing in not existing Input the set of the text file name which is the output of said success text file name sorting section, and a text file name storing dictionary is referred to. Said every one inputted element of a set of a text file name was taken out, and it has the contents output-control section of a document which repeats performing passing the contents output section of a document to all the text file names under said input.

[0021] Thus, an operation of the text file retrieval equipment of constituted this invention is explained using an example, in order to make the understanding easy. First, what was shown in drawing 5 (a) and (b) is used as a WWW text file used as the object which a user searches. In the text file of drawing 5 , the pair of the attribute name and the attribute value of an attribute of an attribute other than a text text is contained. Moreover, the following sentence is used as an example of a user's retrieval sentence.

Retrieval input-statement: "I want to find the information on the hotel which it is possible for a price to take a pet in 150 dollar, and is moreover near San Francisco"

[0022] First, an input statement is changed into a keyword train in the 1st step. Two kinds exist as a class of keyword. The 1st is a natural language expression which refers to an attribute name, and it calls it an attribute name index. The 2nd is a natural language expression which refers to attribute value, and they call it an attribute value index.

[0023]

the generated keyword train:{price (attribute name index of a "price"), $150 dol (attribute value index of a "price"), and a pet (attribute name index of "pet possibility") -- possible (attribute value index of "pet possibility"), San Francisco (attribute value index of a "location"), and a hotel (attribute name index of a "hotel name")

} -- (4) [0024] Next, a redundant part is unified with reference to the order of a list of an attribute name index and an attribute value index. When the attribute name index and attribute value index to the same attribute are located in a line with next doors, the attribute name index is deleted. The upper example of a keyword train is compressed as follows.

the compressed keyword train:{$150 dol (attribute value index of a "price") -- possible (attribute value index of "pet possibility"), San Francisco (attribute value index of a "location"), and a hotel (attribute name index of a "hotel name")

} -- (5) [0025] Next, the extracted keyword train is interpreted. An attribute value index is interpreted as the conditional expression of taking the value which an attribute value index holds as a value of the attribute which it refers to. For example, $150 Dollar (attribute value index of a "price")

**, value = $150" of ""price" attribute

The interpretation to say is carried out.

[0026] When two or more attribute value indexes exist, what carried out the AND of those interpretations serves as the whole conditional expression. In the upper example, it is as follows.

the whole conditional expression -- {-- "-- value = [ ] of the value = $150" and ""pet possibility" attribute of a "price" attribute -- possible --" and "value = San Francisco of a "location" attribute"

} -- (6) [0027] An attribute name index becomes the interpretation "output the value of the attribute which it refers to." In the upper example, it is as follows.

Specific {hotel of a retrieval part" (attribute name index of a "hotel name")

} -- (7) [0028] This semantics serves as the interpretation "output the value of a "hotel name" attribute." When there are two or more attribute name indexes, it becomes the semantics "carry out the sequential output of the attribute name index of these plurality."

[0029] The interpretation of the whole input statement is WWW which satisfies the retrieval conditional expression generated from an attribute value index. What is necessary is to extract the value of the attribute name which chooses the

upper text file, next is specified by the interpretation of an attribute name index out of those text files, and just to display it on a user.

[0030]

[Embodiment of the Invention] When drawing 1 is referred to, the text file retrieval equipment 100 of the gestalt of operation of this invention The text file term document 1, the attribute name storing dictionary 2, and the attribute value storing dictionary 3, With the keyword extraction section 4, the keyword filter section 5, and the contents Banking Inspection Department 6 of a document It consists of the contents Banking Inspection Department 7 of an integrated document, the success text file name sorting section 8, the contents output section 9 of a document, and the contents output-control section 10 of a document, and a display 102 and the Internet 103, such as the input devices 101, such as a keyboard, and a CRT display, are accessed.

[0031] The identifier and the physical location of all the text files used as the candidate for retrieval are stored in the text file term document 1. When the text file used as the candidate for retrieval is described by HTML and XML, a text file is WWW in the world. Distributing to the server is also possible. In that case, the location of a text file is the URL "http://.....". It is described.

[0032] The pair with the natural language expression expressing the attribute name and its attribute name of the attribute tag which exists in the text file used as the candidate for retrieval is registered into the attribute name storing dictionary 2. The most fundamental thing in the natural language expression which refers to a certain attribute name is the attribute name itself. For example, as a natural language expression which refers to the attribute name of a "hotel", it is a "hotel." However, there is an expression which refers to a "hotel" besides it. For example, there are a "stay location" and "an expression which stays." It becomes a pair as these are in the following table 1 and show, and is registered.

[0033]

[Table 1]

ID=000003

[0034] 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value and its attribute value about the attribute value which exists in the text file used as the candidate for retrieval are stored in the attribute value storing dictionary 3. It is [which refers to a certain attribute value] the attribute value itself which is the most fundamental as a natural language expression. There is the "X hotel" itself and there is no other way but to be [a natural language expression which refers to the attribute value of "X hotel"] fastidious. However, in another example, an expression like animal companion being [O.K.] "the lump", pet being [O.K.] "the company", and "dog and cat being good" may also be registered besides "being possible" as a natural language expression showing the attribute value of a "pet" attribute. As shown in the following table 2, data are stored in the attribute value storing dictionary 3 in 3 groups.

[0035]

[Table 2]

ID=000004

[0036] The keyword extraction section 4 is an input unit 101 about the input condition retrieval sentence by natural language expression. If it leads and receives from a user, it will investigate whether with reference to the attribute name storing dictionary 2 and the attribute value storing dictionary 3, there is any expression registered as a natural language expression in it into an input condition retrieval sentence. Only an attribute name is outputted, when it is and it is an attribute name. This output is called an attribute name index. On the other hand, when it is attribute value, the pair of attribute value and a corresponding attribute name is outputted. This output is called an attribute value index. If these are investigated from the head of an input condition retrieval sentence and what matches is found in them, they are outputted to the sequence.

[0037] When the output of the keyword extraction section 4 is investigated sequentially from reception and a head as it is and the attribute name index and attribute value index of the same attribute exist in next doors, the keyword filter section 5 deletes the attribute name index, and outputs the other part transparently as it is.

[0038] The contents Banking Inspection Department 6 of a document receives as an input the character string and attribute value index which are the contents of the text file from the contents Banking Inspection Department 7 of an integrated document. It investigates whether the tag which includes the attribute name in the received attribute value index in the contents character string of the text file received as an input exists, and when it exists, the attribute value which exists by that tag and pair is taken out, investigation etc. is carried out [whether that value is equal to the attribute value in this attribute value index, and], and it is, and success is outputted, and a case carries out a rejected output, when that is not right. The contents Banking Inspection Department 6 of a document has played a kind of subroutine-role which is called from the contents Banking Inspection Department 7 of an integrated document, and operates.

[0039] The contents Banking Inspection Department 7 of an integrated document receives as an input the character string and one or more attribute value indexes which are the contents of the text file from the success text file name sorting section 8. The given attribute value index considers that each is expressing the conditional expression "the value described in the attribute value index must be taken as a value of the attribute described in the attribute value index." The role of the contents Banking Inspection Department 7 of an integrated document is investigating whether the condition being satisfied by finding the attribute expression described by the attribute value index out of the given character string. If all the conditions of one or more attribute value indexes given as an input are satisfied, the value of "success" is outputted, and when that is not right, the value of a "rejection" will be outputted. The contents Banking Inspection Department 6 of a document performs judging whether the character string which is the contents of the text file actually satisfies the conditions of one attribute value index. The contents Banking Inspection Department 7 of an integrated document is performing a kind of loop control which passes every one attribute value index to the contents Banking Inspection Department 6 of a document one by one, when there are two or more attribute value indexes. It is the subroutine-role to which the contents Banking Inspection Department 7 of an integrated document is also called from the success text file name sorting section 8.

[0040] The success text file name sorting section 8 is WWW currently distributed all over the world through the Internet 103 if needed with reference to the text file term document 1. A server is accessed, every one contents of the text file are taken out, and the output of delivery and the contents Banking Inspection Department 7 of an integrated document is received for the part of an attribute value index to the contents Banking Inspection Department 7 of an integrated document among the contents of this document, and the output of the keyword filter section 5. As an output, the value of

"success" or a "rejection" is returned here. This processing is performed to all the files registered into the text file term document 1, and a text file name is outputted to the contents output-control section 10 of a document only to the file whose output of the contents Banking Inspection Department 7 of an integrated document was "success."

[0041] The contents output section 9 of a document inputs a text file name, the contents of this file, and one or more attribute name indexes that are the outputs of the keyword filter section 5. Take out one of the inputted attribute name indexes, and in the contents of the inputted text file Investigate whether a tag including the attribute name in this attribute name index exists, and when it exists It is a display 102 about a pair with the text file name inputted as the value of the attribute value tag corresponding to the attribute name tag. It leads and processing in which nothing is outputted when it does not display and exist in a user is performed to each of all the inputted attribute name indexes. The contents output section 9 of a document is carrying out the role called in subroutine by the contents output-control section 10 of a document. In addition, instead of outputting the pair of the value of an attribute value tag, and a text file name, you may make it display the value of an attribute value tag, and the positional information of a text file, and may make it display the value of an attribute value tag, a text file name, and its positional information.

[0042] The contents output-control section 10 of a document considers the set of the text file name which is the output of the success text file name sorting section 8 as an own input as it is, and the text file term document 1 is referred to. The need is accepted in the contents of the text file under set of the inputted text file name, and it is the Internet 103. It leads and is WWW. Access a server and it takes out one at a time. It repeats performing passing the contents output section 9 of a document to all the text file names under input with the attribute name index generated in a text file name and the keyword filter section 5. That is, when three text file names which passed as an input are received, the contents output section 9 of a document will be called 3 times. In addition, the success text file name sorting section 8 is the Internet 103. It is the Internet 103 by the contents output-control section 10 of a document using the contents, when the contents of the text file which it led and was incorporated from the WWWW server are saved at the magnetic disk drive etc. The count of access can be reduced.

[0043] Drawing 2 and drawing 3 are text file retrieval equipment 100. It is the flow chart which shows the example of processing. Hereafter, actuation of the gestalt of this operation is explained.

[0044] The keyword extraction section 4 is an input unit 101. If it leads and the retrieval input statement by natural language expression is received from a user (step S1), the attribute name storing dictionary 2 and the attribute value storing dictionary 3 will be referred to. Investigate sequentially from the head of a retrieval input statement, and when it is, whether there is any expression registered as a natural language expression in it in a retrieval input statement When it is an attribute name, the attribute name index only containing an attribute name is outputted, and when it is attribute value, the attribute value index containing the pair of attribute value and a corresponding attribute name is outputted (step S2).

[0045] Next, the keyword filter section 5 inspects the list of the index outputted from the keyword extraction section 4, detects the part where the attribute name index and attribute value index of the same attribute are continuing, and deletes the attribute name index of the part (step S3).

[0046] Next, the success text file name sorting section 8 takes out the contents of the document of the text file name, and makes the contents Banking Inspection Department 7 of an integrated document judge delivery and success or failure paying attention to one text file name in the text file term document 1 with all the attribute value indexes outputted from the keyword filter section 5 (step S4).

[0047] The contents Banking Inspection Department 7 of an integrated document makes the contents Banking Inspection Department 6 of a document judge delivery and success or failure for this attribute value index and the contents of the text file paying attention to one of the attribute value indexes passed first, in order to inspect the passed contents of a document (step S5).

[0048] The attribute name tag with which the contents Banking Inspection Department 6 of a document has the attribute name included in the passed attribute value index in the contents of the passed text file exists. And the value of the attribute value tag which is the attribute name tag which existed and pair or [that inspect whether it is in agreement with the attribute value included in the passed attribute value index, and such an attribute name tag does not exist success in being in agreement] -- or even if it exists, when the attribute value is not in agreement, the contents Banking Inspection Department 7 of an integrated document is notified of a rejection (step S6) .

[0049] When, as for the contents Banking Inspection Department 7 of an integrated document, success is notified from the contents Banking Inspection Department 6 of a document (it is YES at step S7), It investigates whether it finish inspecting about all the attribute value indexes notified from the success text file name sorting section 8. When having finished inspecting still, attention is moved to one of NO) and the remaining attribute value indexes at the (step S8, and the contents Banking Inspection Department 6 of a document is made to judge delivery and success or failure for the

attribute value index and contents of the text file (step S9). And when success is declared about all attribute value indexes in the contents Banking Inspection Department 7 of a document (it is YES at step S8), success is notified to the success text file name sorting section 8, the success text file name sorting section 8 uses the text file concerned as a success text file (step S10), and it progresses to step S11. On the other hand, when a rejection is notified from the contents Banking Inspection Department 6 of a document (it is NO at step S7), the contents Banking Inspection Department 7 of an integrated document notifies a rejection to the success text file name sorting section 8, and the success text file name sorting section 8 progresses to step S11.

[0050] After the yes-no decision about one text file finishes, when the unsettled text file remains into the text file term document 1 (it is YES at step S11), the success text file name sorting section 8 moves attention to one of text file names [them] (step S12), and makes the judgment of success or failure like a previous text file.

[0051] If it judges whether the success text file name sorting section 8 had at least one success file (step S13) and there is no one after finishing the yes-no decision to all the text files in the text file term document 1 (it is YES at step S11), the text file corresponding to the retrieval conditions inputted, for example will process displaying the purport which one did not have on a user etc., and will end processing. On the other hand, when at least one success file exists, the text file name of all that success file and all the attribute name indexes outputted from the keyword filter section 5 are notified to the contents output-control section 10 of a document, and the contents output control of a document is made to start (step S14).

[0052] The contents output-control section 10 of a document takes out the contents of a document paying attention to one notified success file name, and the contents output section 9 of a document is made to start processing of delivery and the document concerned with all the notified attribute name indexes (step S15).

[0053] It investigates whether the contents output section 9 of a document has an attribute name tag with the attribute name of the attribute name index in a document paying attention to one notified attribute name index (step S16) (step S17), and if it is (it is YES at step S18), the value and the text file name concerned of the attribute value tag corresponding to the attribute name tag will be displayed on a display 102 (step S19). If there is nothing (it is NO at step S18), step S19 will be skipped. Next, if the contents output section 9 of a document investigated whether the unsettled attribute name index would remain on the notified attribute name index (step S20) and it remains in it, it will move attention to one of them (step S21), and will repeat the processing returned and mentioned above to step S17.

[0054] After finishing the processing about all attribute name indexes in which the contents output section 9 of a document was notified (it is NO at step S20), the contents output-control section 10 of a document When investigating whether the unsettled thing remains to the text file notified from the success text file name sorting section 8 (step S22) and remaining in it, move attention to one of them, and the contents of a document of the text file name are taken out. The contents output section 9 of a document is passed, and is made to process with all the attribute name indexes notified from the success text file name sorting section 8 (step S23). It becomes processing termination after the processing about all success files finishes (it is YES at step S22).

[0055]

[Example] Suppose that three text file names, "a file 1", "a file 2", and "a file 3", and URL of those are registered into the text file term document 1 so that it may illustrate to drawing 1 . Moreover, the contents of the file 1 show drawing 5 (a), and the contents of the file 2 presuppose that it is what is shown in drawing 5 (b). These files 1 and 2 are XML. It is described and the attribute and attribute value other than a text text are included. That is, the pair of the tag which the attribute name expressing the semantics written into the document of a <hotel> X hotel <<location> San Francisco </location> <price> $150 </price> <pet> possible < / pet> attached, and the value of the attribute is contained in the file 1. Similarly, a tag called a <hotel> Z hotel <<location> Seattle </location> <price> $180 </price> <pet> improper < / pet> is contained also in the file 2.

[0056] Moreover, the pair of an attribute name which is illustrated to drawing 1 , and its natural language expression shall be stored in the attribute name storing dictionary 2 in advance, and 3 groups of attribute value and a natural language expression which are illustrated to drawing 1 , and an attribute name shall be stored in the attribute value storing dictionary 3 in advance. In addition, if all prices are registered into the attribute value storing dictionary 3 with the actual value, since a number of registration will increase, you may make it register using a variable. That is, XXX If it registers as it is shown in the following table 3, when considering as the numeric value of arbitration, and the keyword extraction section 4 has a "dollar" behind the numeric value of arbitration, it will be the natural language expression XXX. Let what judged that a dollar existed and attached $ to the head of the actual value which existed be attribute value.

[Table 3]

ID=000005

[0057] Actuation of this example is explained to an example for the case where a user inputs the retrieval input statement by the following natural language by such premise.
Retrieval input-statement: "I want to find the information on the hotel which it is possible for a price to take a pet in 150 dollar, and is moreover near San Francisco"
[0058] The keyword extraction section 4 will change a retrieval input statement into a keyword train as follows with reference to the attribute name storing dictionary 2 and the attribute value storing dictionary 3, if the retrieval input statement from a user is received.
[0059] First, since the natural language expression "a price" of the head of a retrieval input statement exists in the attribute name storing dictionary 2, the attribute name "a price" registered into it and a pair by becoming is outputted as an attribute name index. Next, since a natural language expression "150 Dollar" exists in the attribute value storing dictionary 3, the pair of the attribute value "$150" and the attribute name "the price" which are registered in it and 3 groups is outputted as an attribute value index. Next, since a natural language expression "a pet" exists in the attribute name storing dictionary 2, the attribute name "a pet" registered into it and a pair by becoming is outputted as an attribute name index. Next, since a natural language expression "possible" exists in the attribute value storing dictionary 3, the pair of the attribute value "possible" and the attribute name "the pet" which are registered in it and 3 groups is outputted as an attribute value index. Next, since a natural language expression "San Francisco" exists in the attribute value storing dictionary 3, the pair of the attribute value "San Francisco" and the attribute name "the location" which are registered in it and 3 groups is outputted as an attribute value index. Next, since a "hotel" exists in the attribute name storing dictionary 2, the attribute name "a hotel" registered into it and a pair by becoming is outputted as an attribute name index. Into a retrieval input statement, other natural language expressions which match the natural language expression registered into the attribute name storing dictionary 2 and the attribute value storing dictionary 3 cannot be found. Therefore, the following keyword trains are outputted sequentially from a top.
[0060] Attribute name index (attribute name "a price")
Attribute value index (attribute value "$150", attribute name "a price")
Attribute name index (attribute name "a pet")
Attribute value index (attribute value "possible", attribute name "a pet")
Attribute value index (attribute value " "San Francisco", attribute name "a location")
Attribute name index (attribute name "a hotel")
[0061] Next, the keyword filter section 5 unifies a redundant part with reference to the order of a list of an attribute name index and an attribute value index. Since it has ranked with next doors by the attribute name "a price" with same attribute name index (attribute name "price") and attribute value index (attribute value "$150", attribute name "price") in the case of the upper keyword train, an attribute name index (attribute name "a price") is deleted. Moreover, since it has ranked with next doors by the attribute name "a pet" with same attribute name index (attribute name "pet") and attribute value index (attribute value "possible", attribute name "pet"), an attribute name index (attribute name "a pet") is deleted. Since the attribute name index which should otherwise be deleted does not exist, finally the above-mentioned keyword train is compressed as follows.
[0062] (a) Attribute value index (attribute value "$150", attribute name "a price")
(b) Attribute value index (attribute value "possible", attribute name "a pet")
(c) Attribute value index (attribute value " "San Francisco", attribute name "a location")
(d) Attribute name index (attribute name "a hotel")
[0063] next, the success text file name sorting section 8 -- the contents of a document of the file 1 in the text file term document 1 -- the URL -- reliance -- the Internet 103 from the server which leads and corresponds -- acquiring -- the contents of a document, and the above-mentioned attribute value index (a) - (c) The contents Banking Inspection Department 7 of an integrated document is passed.
[0064] The contents Banking Inspection Department 7 of an integrated document is the contents of a document of a file 1, and one attribute value index (a). The contents Banking Inspection Department 6 of a document is passed.
[0065] The contents Banking Inspection Department 6 of a document is an attribute value index (a) in the contents of a

document of a file 1. It investigates whether the tag of an inner attribute name "a price" exists. drawing 5 -- (-- a --) -- a file -- one -- a case -- corresponding -- a tag -- < -- a price -- > -- $ -- 150 -- < -- /-- a price -- > -- it is -- since -- the -- attribute value -- "-- $ -- 150 -- " -- having received -- attribute value -- an index -- (-- a --) It investigates whether it is in agreement with inner attribute value "$150." In the present example, since it is in agreement, success is returned to the contents Banking Inspection Department 7 of an integrated document.

[0066] The contents Banking Inspection Department 7 of an integrated document is the contents of a document of a file 1, and the following attribute value index (b). The contents Banking Inspection Department 6 of a document is passed.

[0067] The contents Banking Inspection Department 6 of a document is an attribute value index (b) in the contents of a document of a file 1. It investigates whether the tag of an inner attribute name "a pet" exists. drawing 5 -- (-- a --) -- a file -- one -- a case -- corresponding -- a tag -- < -- a pet -- > -- possible -- < -- /-- a pet -- > -- it is -- since -- the -- attribute value -- "-- possible -- " -- having received -- attribute value -- an index -- (-- b --) It investigates whether it is in agreement with inner attribute value "possible." In the present example, since it is in agreement, success is returned to the contents Banking Inspection Department 7 of an integrated document.

[0068] The contents Banking Inspection Department 7 of an integrated document is the contents of a document of a file 1, and the following attribute value index (c). The contents Banking Inspection Department 6 of a document is passed.

[0069] The contents Banking Inspection Department 6 of a document is an attribute value index (c) in the contents of a document of a file 1. It investigates whether the tag of an inner attribute name "a location" exists. drawing 5 -- (-- a --) -- a file -- one -- a case -- corresponding -- a tag -- < -- a location -- > -- San Francisco -- < -- /-- a location -- > -- it is -- since -- the -- attribute value -- "-- San Francisco -- " -- having received -- attribute value -- an index -- (-- c --) It investigates whether it is in agreement with inner attribute value "San Francisco." In the present example, since it is in agreement, success is returned to the contents Banking Inspection Department 7 of an integrated document.

[0070] Since the result of success was obtained by all attribute value indexes about the file 1, the contents Banking Inspection Department 7 of an integrated document notifies success to the success text file name sorting section 8, and the success text file name sorting section 8 considers a file 1 as a success file.

[0071] Next, the success text file name sorting section 8 is the Internet 103 to reliance about the URL in the contents of a document of the file 2 stored in the text file term document 1. It leads, and incorporates from the corresponding server and success or failure is judged using the contents Banking Inspection Department 7 of an integrated document like the previous file 1. In this case, since a pet attribute and a location attribute are not satisfied, a file 2 serves as a rejection. Similarly, the judgment of success or failure is performed also about the remaining file 3. Here, a file 3 is also judged to be a rejection and a success file presupposes that it was it only a file 1.

[0072] Next, the success text file name sorting section 8 is an attribute name index (d) about a file name 1 as a success file name. The contents output-control section 10 of a document is passed.

[0073] The contents output-control section 10 of a document acquires URL of a file name 1 from the text file term document 1, and is the Internet 103 to reliance about it. The upper server is accessed, the contents of a document of a file name 1 are acquired, and it is an attribute name index (d). The contents output section 9 of a document is passed.

[0074] The contents output section 9 of a document is an attribute name index (d) in the contents of a document of a file 1. It investigates whether an attribute tag with an attribute name "a hotel" exists. since the corresponding tag <hotel> X hotel <a /hotel> exists in the case of the file 1 of drawing 5 (a) -- the attribute value "X hotel" -- taking out -- a file name 1 -- display 102 It displays.

[0075] Drawing 4 is the block diagram showing the example of a hardware configuration of the text file retrieval equipment of this invention. The text file retrieval equipment of this example A processor 200 (CPU), The auxiliary storage units 201, such as a magnetic disk drive, and the interface 202 of those, The memory 203, such as RAM, and the interface 204 between the Internet 103, The record medium 205 which can machine read CD-ROM, semiconductor memory, etc., It consists of buses 209 which connect between the interface 206, an input unit 101, its interface 207, a display 102, its interface 208, and CPU200, memory 203 and an interface 202,204,206-208.

[0076] The program for text file retrieval is recorded on the record medium 205, and the text file term document 1, the attribute name storing dictionary 2, and the attribute value storing dictionary 3 of drawing 1 are loaded on memory 203 or an auxiliary storage unit 201 by installing this program through an interface 206. Moreover, this program operates CPU200 by controlling actuation of CPU200 as the keyword extraction section 4 of drawing 1 , the keyword filter section 5, the contents Banking Inspection Department 6 of a document, the contents Banking Inspection Department 7 of an integrated document, the success text file name sorting section 8, the contents output section 9 of a document, and the contents output-control section 10 of a document.

[0077] The gestalt of the above operation is WWW about this invention. Although applied to retrieval of the upper homepage, it is applicable also to the retrieval to a database like the technique indicated to JP,5-67136,A.

[0078]
[Effect of the Invention] According to this invention, the following effectiveness is acquired as explained above.
[0079] The redundancy of the reply to the retrieval inquiry by natural language can be abolished as much as possible. The reason is that it has the keyword filter section which deletes an attribute name index when the index train extracted in the keyword extraction section is investigated sequentially from the head and the attribute name index and attribute value index of the same attribute exist in next doors.
[0080] WWW The retrieval inquiry by natural language becomes possible also to a homepage. The reason is XML. It is because the attribute value which a user desires can be taken out out of a text file with the attribute name and attribute value which interpret the retrieval conditions by the natural language which was made to build in the tag of the attribute name expressing the semantics, and attribute value, and the user inputted into the text file like, and suit. WWW by the natural language interface which receives the input of various forms, such as a fragment of a right natural language expression, a nongrammatical expression, and a natural language sentence, and a keyword train, grammatically, and interprets systematically by this A document-retrieval system is realizable.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

TECHNICAL FIELD

[Field of the Invention] Especially this invention relates to the text file retrieval equipment which made the retrieval inquiry by natural language possible about text file retrieval equipment.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

PRIOR ART

[Description of the Prior Art] When it is going to make a precision express a retrieval intention of a user more in information retrieval generally, the method of making it express with natural language like Japanese or English is effective. The system which performs retrieval to a database with natural language already exists, and is called the natural language interface (bibliography: DIBEROPPINGUANACHURARURANGEJI interface two complex data and G . G . "Developing a Natural Language Interface to Complex Data" besides Hendrix, ACM Trans.on Database Systems, 1978.).

[0003] The conventional natural language interface interprets the retrieval inquiry by a user's natural language, changes the inquiry into the retrieval type of the query language (SQL) of a database, and shows a user the retrieval result to which the retrieval type was returned by database system from delivery and database system.

[0004] However, although commercialization already started and the conventional natural language interface left for 20 years or more, it has not reached the level of practical use yet. Since the functor in which a natural language interface system cannot interpret a free inquiry of a user, but the system permits one of the reason of the, and a vocabulary are limited clearly, a user is for having to memorize what kind of expression can be used after all. Therefore, a natural language interface is not different from a complicated command system. That is, it was a problem to say that the conventional natural language interface cannot receive a user's free expression (bibliography: DIBEROPPINGUANACHURARURANGEJI interface two complex data and G . G . "Developing a Natural Language Interface to Complex Data" besides Hendrix, ACM Trans.on Database Systems, 1978.). It was especially a problem that it is not allowed although he wants to use not the natural language with a perfect user about an easy question but a brief expression. As an example of a brief expression, there are a part of expression by the keyword train, nongrammatical expression or natural language sentence, and **.

[0005] Then, these people proposed the "natural language interpretation approach" of realizing retrieval by the natural language of a brief expression, in previous patent application (JP,5-67136,A). This makes applicable to retrieval the table of the database which is the assembly of the group of the attribute name and the attribute value of an attribute of an attribute. Each word in the inquiry sentence by natural language is classified into an attribute name, attribute value, and others. Save the word classified into the attribute name as a response attribute name, and the word classified into attribute value makes a group the attribute value and the attribute name corresponding to it, and it saves it as condition attribute value ****. When all the groups of the attribute value in this saved condition attribute value **** and an attribute name exist in a table, the attribute value in said table corresponding to each attribute name saved as said response attribute name is outputted as a reply to an inquiry. In addition, there is the "natural language interpretation approach" which starts JP,5-242147,A for which these people applied previously too as a conventional technique similar to this.

[0006] On the other hand, it follows on expansion of use of World Wide Web (WWW) in recently, and is WWW. The importance of the upper retrieval technique is increasing. WWW A user is WWW. The typical tool used when carrying out information retrieval in a top is a search engine. As this example, Altavista, Infoseek, Lycos, etc. are famous. However, in a search engine, since it is the format of the retrieval which combines a keyword, the intention of retrieval of a user is not reflected directly in many cases. For example, a price is 150 when looking for the information about a hotel. It is possible to take a pet in a dollar, and it is impossible to express such a retrieval intention only along a keyword to find the homepage of the hotel which is moreover near San Francisco. Temporarily, it is (1). It expresses by the formula. 150 A dollar, San Francisco, a hotel, and a pet are possible. -- If (1) and a keyword are put in order, the list of very a lot of homepages will be outputted.

[0007] A natural language interface system is introduced and it is WWW. If a homepage can be searched, precise retrieval conditions like the upper example will be expressed obediently, and only a suitable homepage can be searched.

However, WWW The example which applied the natural language interface to retrieval of the upper homepage is not found.

[0008] in addition, WWW although there be "information retrieval equipment" indicated by JP,10-40262,A as other examples of the retrieval technique over a homepage, since this aim at the information retrieval suitable for the sensibility of the user who do not have the retrieval conditions for [clear] retrieval by enable retrieval which made sensibility expression data the keyword, it be fit for retrieval by precise retrieval conditions like the upper example.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

EFFECT OF THE INVENTION

[Effect of the Invention] According to this invention, the following effectiveness is acquired as explained above.
[0079] The redundancy of the reply to the retrieval inquiry by natural language can be abolished as much as possible. The reason is that it has the keyword filter section which deletes an attribute name index when the index train extracted in the keyword extraction section is investigated sequentially from the head and the attribute name index and attribute value index of the same attribute exist in next doors.
[0080] WWW The retrieval inquiry by natural language becomes possible also to a homepage. The reason is XML. It is because the attribute value which a user desires can be taken out out of a text file with the attribute name and attribute value which interpret the retrieval conditions by the natural language which was made to build in the tag of the attribute name expressing the semantics, and attribute value, and the user inputted into the text file like, and suit. WWW by the natural language interface which receives the input of various forms, such as a fragment of a right natural language expression, a nongrammatical expression, and a natural language sentence, and a keyword train, grammatically, and interprets systematically by this A document-retrieval system is realizable.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

TECHNICAL PROBLEM

[Problem(s) to be Solved by the Invention] Although the technique indicated by JP,5-67136,A mentioned above can serve as an effective means when putting a natural language interface system in practical use since it can interpret the inquiry by natural language by the simple approach, the technical problem which should still be solved is left behind. Since it is surely treated as a response attribute name and is included during the reply to an inquiry when a certain attribute name exists in the inquiry sentence by natural language, it is that a reply may become redundancy.
[0010] For example, a table with "an attribute name = title, its attribute value = No Longer Human, an attribute name = author, and its attribute value = Dazaiosamu" is received. When the inquiry "the author of title = No Longer Human is ?" is performed, first Subsequently, "No Longer Human" is judged to be attribute value, a group with the "title" which is attribute value "No Longer Human" and its attribute name is saved as condition attribute value ****, a "title" is judged to be an attribute name, and it is saved as a response attribute name, and he is saved [an "author" is judged to be an attribute name and] as a response attribute name. And said table which has the attribute value in condition attribute value **** "No Longer Human" and a group with an attribute name "a title" is searched, and the attribute value "No Longer Human" and "Dazaiosamu" corresponding to a response attribute name "a title" and an "author" are searched and outputted from the table. That is, the part and reply which are outputting "No Longer Human" are redundancy.
[0011] Then, the purpose of this invention is to abolish the redundancy of the reply to the retrieval inquiry by natural language as much as possible.
[0012] Moreover, other purposes of this invention are WWW. It is in making the retrieval inquiry by natural language possible also to a homepage.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

MEANS

[Means for Solving the Problem] (1) In the technique indicated by JP,5-67136,A in which the 1st carried out invention ****, although the user asked the attribute value corresponding to it and the reason the reply is redundancy has described it in a sentence when an attribute name exists in the inquiry sentence by natural language, it is because it is treating as a response attribute name uniformly. Then, when the attribute name and attribute value of the same attribute appear in next doors, he is trying not to include the attribute name in a response attribute name in this invention paying attention to the point in the inclination which a user adjoins them mutually and is described when describing a certain attribute name and the attribute value corresponding to it, in order to specify retrieval conditions in the inquiry sentence by natural language. The text file which contains the pair of the attribute name and the attribute value of an attribute of an attribute is more specifically used as the text file for retrieval. In the text file retrieval equipment with which the part which suits the retrieval conditions which the user specified with natural language from the text file for retrieval is searched Investigate the retrieval demand sentence expressed with natural language sequentially from a head, and the attribute name is outputted as an attribute name index to the natural language expression expressing an attribute name. The keyword extraction section which performs outputting the pair of the attribute value and attribute name as an attribute value index to the natural language expression expressing attribute value one by one, Only when the output of said keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors The keyword filter section which deletes said attribute name index and outputs the other part as it is, It investigates whether the pair of the attribute name of all attribute value indexes and attribute value which were outputted from said keyword filter section into the text file for retrieval is built in. When built, it has a retrieval means to search and output the attribute value corresponding to the attribute name of the attribute name index outputted from said keyword filter section from the text file for retrieval.
[0014] Furthermore, in order to enable it to judge correctly which natural language expression which natural language expression in a retrieval demand sentence expresses an attribute name, and is expressing attribute value The attribute name storing dictionary which stores the pair with the natural language expression expressing an attribute name and its attribute name about the attribute name which exists in the text file for retrieval, It has the attribute value storing dictionary which stores 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value, and its attribute value about the attribute value which exists in the text file for retrieval. Said keyword extraction section investigates the retrieval demand sentence expressed with natural language sequentially from a head, and refers to an attribute name storing dictionary. If the natural language expression expressing an attribute name is included, will output the attribute name which is the natural language expression and pair as an attribute name index, and an attribute value storing dictionary will be referred to. If the natural language expression expressing attribute value is included, it has the configuration which outputs a pair of set with the natural language expression, the attribute value which is 3 groups, and an attribute name as an attribute value index.
[0015] Thus, if it is in the text file retrieval equipment of constituted this invention When a user inputs the retrieval demand sentence expressed with natural language, first the keyword extraction section Investigate a retrieval demand sentence sequentially from a head, and the attribute name is outputted as an attribute name index to the natural language expression expressing an attribute name. To the natural language expression expressing attribute value, the pair of the attribute value and attribute name is outputted as an attribute value index. Only when the keyword filter section inputs the output of the keyword extraction section, and investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors, an attribute name index is deleted. Subsequently, subsequently It investigates whether the pair of the attribute name of all the attribute value indexes with which the retrieval means was outputted from the keyword filter section into the text file for retrieval, and attribute value is built in. When built, it has prevented the reply to a user becoming redundancy by searching and outputting the attribute value

corresponding to the attribute name of the attribute name index outputted from the keyword filter section from the text file for retrieval.
[0016] (2) 2nd invention WWW For applying a natural language interface to retrieval of the upper homepage, the difficult reason is WWW. The contents of the homepage are the files which consist of the texts and drawings which were drawn with natural language, and it is not being the set of an attribute name and attribute value like a database. That is, the target database [interface system / which is the former / natural language] is supposing there is a database of the hotel which comes out in the example of the term of a Prior art since it was the set of an attribute name and attribute value (identifier: X hotel, price:$150, pet:possibility of, location:San Francisco),
(Identifier: Y hotel, price:$200, pet:improper, location:Los Angeles)
(Identifier: Z hotel, price:$180, pet:improper, location:Seattle)
** -- it stores with a gestalt [like] -- having -- **** -- an inquiry of a user -- (2) it expresses by the formula -- as -- "-- price =$150 and pet = -- possible and location = San Francisco"
-- (2) It was [like] convertible. This is SQL. It is changed into language as it is. However, WWW Since the information on such [usually] an attribute name and attribute value is not contained, a homepage is SQL. It is not convertible for the retrieval type which can be expressed by the formula. That is, the conventional WWW Homepage creation language is HTML (bibliography: the homepage of a World-Wide-Web consortium, URL http://www.w3.org). In HTML, the group of the attribute name for expressing the configuration and attribute value is built in in the text file. For example, it is the example of an HTML file which is shown in drawing 6 . here -- < -- what it is the attribute tag which was surrounded by > and used independently (example : <B>) Initiation tag (example : <TR>) Termination tag (example : </TR>) There are some which are used by the pair. The description of a HTML tag is limited to it defining the expression of the appearance in a text file. For example, the tag made a table-expression is <TABLE> and the tag showing line feed is expressed by <P>. It is WWW about such an HTML file. If it is made to read into a browser, it will become a gestalt as shown in drawing 7 , and will be indicated by the output at a user. However, HTML cannot define the tag for expressing the semantics in a document.
[0017] So, at this invention, it is WWW. The group of the attribute name and attribute value expressing the semantics in the text file is made to build in a text file. XML (Extensible Markup Language) specifically extended so that the contents of the document could be expressed by a pair of set of an attribute tag and its attribute value in a file describes a document (bibliography: the homepage of a World-Wide-Web consortium, "EKUSUTENSHIBURU markup language 1.0"http://www.w3.org/TR/PR-xml -971208). . XML WWW The specification was proposed by the World-Wide-Web consortium (bibliography: the homepage of a World-Wide-Web consortium, URL http://www.w3.org) which is the engine which decides a criterion in December, 1997. XML By the described document, a machine becomes legible about the contents of the document and a content addressed retrieval becomes possible. Then, that is used in this invention and it is WWW. The retrieval inquiry by natural language is realized to a homepage.
[0018] The text file which contains the pair of the tag which the attribute name which specifically expresses the semantics written into the document attached, and the attribute value of the attribute is used as the text file for retrieval. In the text file retrieval equipment with which the part which suits the retrieval conditions which the user specified with natural language from the text file for retrieval is searched The attribute name storing dictionary which stores a pair of set with the natural language expression expressing an attribute name and its attribute name about the attribute name which exists in the text file for retrieval, The attribute value storing dictionary which stores the set of 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value, and its attribute value about the attribute value which exists in the text file for retrieval, Investigate the retrieval demand sentence expressed with natural language sequentially from a head, and an attribute name storing dictionary is referred to. If the natural language expression expressing an attribute name is included, will output the attribute name which is the natural language expression and pair as an attribute name index, and an attribute value storing dictionary will be referred to. The keyword extraction section which will output a pair of set with the natural language expression, the attribute value which is 3 groups, and an attribute name as an attribute value index if the natural language expression expressing attribute value is included, Only when the output of the keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors The keyword filter section which deletes said attribute name index and outputs the other part as it is, It investigates whether the tag corresponding to the pair of the attribute name of all attribute value indexes and attribute value which were outputted from said keyword filter section into the text file for retrieval is built in. When built, it has a retrieval means to search and output the attribute value of a tag with the attribute name of the attribute name index outputted from said keyword filter section from the text file for retrieval.
[0019] Thus, if it is in the text file retrieval equipment of constituted this invention When a user inputs the retrieval

demand sentence specified with natural language, the keyword extraction section If the natural language expression which investigates a retrieval demand sentence sequentially from a head, and expresses an attribute name is included The attribute name is outputted as an attribute name index, and if the natural language expression expressing attribute value is included, a pair of set with the attribute value and attribute name will be outputted as an attribute value index. Subsequently Only when the keyword filter section inputs the output of the keyword extraction section, and investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors, said attribute name index is deleted. Subsequently It investigates whether the tag corresponding to the pair of the attribute name of all the attribute value indexes with which the retrieval means was outputted from the keyword filter section into the text file for retrieval, and attribute value is built in. When built, the attribute value of a tag with the attribute name of the attribute name index outputted from the keyword filter section is searched and outputted from the text file for retrieval.

[0020] Moreover, in order to sort out only the text file which fulfills the retrieval conditions which the user inputted with natural language from from among the text files of a large number registered beforehand and to enable it to display the part which the user in it still needs on a user From the set of the text file to build [two or more] in, the pair of the tag which the attribute name expressing the semantics written into the document attached, and the value of the attribute In the text file retrieval equipment which chooses the text file with which are satisfied of the retrieval conditions which the user specified with natural language, and displays the suiting part The text file term document which stores the identifier and existence location of all the text files used as the candidate for retrieval, The attribute name storing dictionary which stores a pair of set with the natural language expression expressing an attribute name and its attribute name about the attribute name which exists in the text file used as the candidate for retrieval, The attribute value storing dictionary which stores the set of 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value, and its attribute value about the attribute value which exists in the text file used as the candidate for retrieval, If a user inputs the retrieval demand sentence expressed with natural language, he will investigate said input statement sequentially from a head, and will refer to an attribute name storing dictionary. If the natural language expression expressing an attribute name is included, will output the attribute name which is the natural language expression and pair as an attribute name index, and an attribute value storing dictionary will be referred to. The keyword extraction section which performs one by one outputting a pair of set with the natural language expression, the attribute value which is 3 groups, and an attribute name as an attribute value index if the natural language expression expressing attribute value is included, Only when the output of the keyword extraction section is inputted, it investigates sequentially from a head and the attribute name index and attribute value index of the same attribute exist in next doors The keyword filter section which deletes said attribute name index and outputs the other part as it is, When the contents and the attribute value index of a text file are inputted, in the contents of said text file Investigate whether a tag including the attribute name in said attribute value index exists, and when it exists Take out the attribute value which exists by the tag and pair, and whether the value is equal to the attribute value in said attribute value index investigation and when equal With the contents Banking Inspection Department of a document which outputs success, and does a rejected output when that is not right If the contents of the text file and one or more attribute value indexes are inputted Take out one at a time from said attribute value index, and the contents and said every one taken-out attribute value index of said text file are passed to the contents Banking Inspection Department of a document. To all attribute value indexes, when the output is success With the contents Banking Inspection Department of an integrated document which outputs success, and outputs a rejection when that is not right Every one contents of the text file are taken out with reference to a text file term document. The contents of said document, and the part of the attribute value index of the outputs of the keyword filter section to the contents Banking Inspection Department of an integrated document Delivery, The success text file name sorting section to which it performs receiving the output of said contents Banking Inspection Department of an integrated document to said all text files that it took out one [at a time], and said output outputs only the identifier of the text file of success, If the contents of a text file name and said text file name and the attribute name index which is the output of the keyword filter section are inputted When one of said attribute name indexes is taken out, and it investigates whether a tag including said taken-out attribute name exists and exists in the contents of said given text file with the contents output section of a document which said inputted attribute name index is alike, respectively, receives, and performs displaying the value and said inputted text file name of a tag of the attribute name on a user, and outputting nothing in not existing Input the set of the text file name which is the output of said success text file name sorting section, and a text file name storing dictionary is referred to. Said every one inputted element of a set of a text file name was taken out, and it has the contents output-control section of a document which repeats performing passing the contents output section of a document to all the text file names under said input.

[0021] Thus, an operation of the text file retrieval equipment of constituted this invention is explained using an example,

in order to make the understanding easy. First, what was shown in drawing 5 (a) and (b) is used as a WWW text file used as the object which a user searches. In the text file of drawing 5 , the pair of the attribute name and the attribute value of an attribute of an attribute other than a text text is contained. Moreover, the following sentence is used as an example of a user's retrieval sentence.
Retrieval input-statement: "I want to find the information on the hotel which it is possible for a price to take a pet in 150 dollar, and is moreover near San Francisco"
[0022] First, an input statement is changed into a keyword train in the 1st step. Two kinds exist as a class of keyword. The 1st is a natural language expression which refers to an attribute name, and it calls it an attribute name index. The 2nd is a natural language expression which refers to attribute value, and they call it an attribute value index.
[0023]
the generated keyword train:{price (attribute name index of a "price"), $150 dol (attribute value index of a "price"), and a pet (attribute name index of "pet possibility") -- possible (attribute value index of "pet possibility"), San Francisco (attribute value index of a "location"), and a hotel (attribute name index of a "hotel name")
} -- (4) [0024] Next, a redundant part is unified with reference to the order of a list of an attribute name index and an attribute value index. When the attribute name index and attribute value index to the same attribute are located in a line with next doors, the attribute name index is deleted. The upper example of a keyword train is compressed as follows.
the compressed keyword train:{$150 dol (attribute value index of a "price") -- possible (attribute value index of "pet possibility"), San Francisco (attribute value index of a "location"), and a hotel (attribute name index of a "hotel name")
} -- (5) [0025] Next, the extracted keyword train is interpreted. An attribute value index is interpreted as the conditional expression of taking the value which an attribute value index holds as a value of the attribute which it refers to. For example, $150 Dollar (attribute value index of a "price")
**, value = $150" of ""price" attribute
The interpretation to say is carried out.
[0026] When two or more attribute value indexes exist, what carried out the AND of those interpretations serves as the whole conditional expression. In the upper example, it is as follows.
the whole conditional expression -- {-- "-- value = [ ] of the value = $150" and ""pet possibility" attribute of a "price" attribute -- possible --" and "value = San Francisco of a "location" attribute"
} -- (6) [0027] An attribute name index becomes the interpretation "output the value of the attribute which it refers to." In the upper example, it is as follows.
Specific {hotel of a retrieval part" (attribute name index of a "hotel name")
} -- (7) [0028] This semantics serves as the interpretation "output the value of a "hotel name" attribute." When there are two or more attribute name indexes, it becomes the semantics "carry out the sequential output of the attribute name index of these plurality."
[0029] The interpretation of the whole input statement is WWW which satisfies the retrieval conditional expression generated from an attribute value index. What is necessary is to extract the value of the attribute name which chooses the upper text file, next is specified by the interpretation of an attribute name index out of those text files, and just to display it on a user.
[0030]
[Embodiment of the Invention] When drawing 1 is referred to, the text file retrieval equipment 100 of the gestalt of operation of this invention The text file term document 1, the attribute name storing dictionary 2, and the attribute value storing dictionary 3, With the keyword extraction section 4, the keyword filter section 5, and the contents Banking Inspection Department 6 of a document It consists of the contents Banking Inspection Department 7 of an integrated document, the success text file name sorting section 8, the contents output section 9 of a document, and the contents output-control section 10 of a document, and a display 102 and the Internet 103, such as the input devices 101, such as a keyboard, and a CRT display, are accessed.
[0031] The identifier and the physical location of all the text files used as the candidate for retrieval are stored in the text file term document 1. When the text file used as the candidate for retrieval is described by HTML and XML, a text file is WWW in the world. Distributing to the server is also possible. In that case, the location of a text file is the URL "http://.....". It is described.
[0032] The pair with the natural language expression expressing the attribute name and its attribute name of the attribute tag which exists in the text file used as the candidate for retrieval is registered into the attribute name storing dictionary 2. The most fundamental thing in the natural language expression which refers to a certain attribute name is the attribute name itself. For example, as a natural language expression which refers to the attribute name of a "hotel", it is a "hotel." However, there is an expression which refers to a "hotel" besides it. For example, there are a "stay location" and "an

expression which stays." It becomes a pair as these are in the following table 1 and show, and is registered.
[0033]
[Table 1]

ID=000003

[0034] 3 groups with the natural language expression which expresses the attribute name corresponding to attribute value and its attribute value and its attribute value about the attribute value which exists in the text file used as the candidate for retrieval are stored in the attribute value storing dictionary 3. It is [which refers to a certain attribute value] the attribute value itself which is the most fundamental as a natural language expression. There is the "X hotel" itself and there is no other way but to be [a natural language expression which refers to the attribute value of "X hotel"] fastidious. However, in another example, an expression like animal companion being [O.K.] "the lump", pet being [O.K.] "the company", and "dog and cat being good" may also be registered besides "being possible" as a natural language expression showing the attribute value of a "pet" attribute. As shown in the following table 2, data are stored in the attribute value storing dictionary 3 in 3 groups.
[0035]
[Table 2]

ID=000004

[0036] The keyword extraction section 4 is an input unit 101 about the input condition retrieval sentence by natural language expression. If it leads and receives from a user, it will investigate whether with reference to the attribute name storing dictionary 2 and the attribute value storing dictionary 3, there is any expression registered as a natural language expression in it into an input condition retrieval sentence. Only an attribute name is outputted, when it is and it is an attribute name. This output is called an attribute name index. On the other hand, when it is attribute value, the pair of attribute value and a corresponding attribute name is outputted. This output is called an attribute value index. If these are investigated from the head of an input condition retrieval sentence and what matches is found in them, they are outputted to the sequence.
[0037] When the output of the keyword extraction section 4 is investigated sequentially from reception and a head as it is and the attribute name index and attribute value index of the same attribute exist in next doors, the keyword filter section 5 deletes the attribute name index, and outputs the other part transparently as it is.
[0038] The contents Banking Inspection Department 6 of a document receives as an input the character string and attribute value index which are the contents of the text file from the contents Banking Inspection Department 7 of an

integrated document. It investigates whether the tag which includes the attribute name in the received attribute value index in the contents character string of the text file received as an input exists, and when it exists, the attribute value which exists by that tag and pair is taken out, investigation etc. is carried out [whether that value is equal to the attribute value in this attribute value index, and], and it is, and success is outputted, and a case carries out a rejected output, when that is not right. The contents Banking Inspection Department 6 of a document has played a kind of subroutine-role which is called from the contents Banking Inspection Department 7 of an integrated document, and operates.

[0039] The contents Banking Inspection Department 7 of an integrated document receives as an input the character string and one or more attribute value indexes which are the contents of the text file from the success text file name sorting section 8. The given attribute value index considers that each is expressing the conditional expression "the value described in the attribute value index must be taken as a value of the attribute described in the attribute value index." The role of the contents Banking Inspection Department 7 of an integrated document is investigating whether the condition being satisfied by finding the attribute expression described by the attribute value index out of the given character string. If all the conditions of one or more attribute value indexes given as an input are satisfied, the value of "success" is outputted, and when that is not right, the value of a "rejection" will be outputted. The contents Banking Inspection Department 6 of a document performs judging whether the character string which is the contents of the text file actually satisfies the conditions of one attribute value index. The contents Banking Inspection Department 7 of an integrated document is performing a kind of loop control which passes every one attribute value index to the contents Banking Inspection Department 6 of a document one by one, when there are two or more attribute value indexes. It is the subroutine-role to which the contents Banking Inspection Department 7 of an integrated document is also called from the success text file name sorting section 8.

[0040] The success text file name sorting section 8 is WWW currently distributed all over the world through the Internet 103 if needed with reference to the text file term document 1. A server is accessed, every one contents of the text file are taken out, and the output of delivery and the contents Banking Inspection Department 7 of an integrated document is received for the part of an attribute value index to the contents Banking Inspection Department 7 of an integrated document among the contents of this document, and the output of the keyword filter section 5. As an output, the value of "success" or a "rejection" is returned here. This processing is performed to all the files registered into the text file term document 1, and a text file name is outputted to the contents output-control section 10 of a document only to the file whose output of the contents Banking Inspection Department 7 of an integrated document was "success."

[0041] The contents output section 9 of a document inputs a text file name, the contents of this file, and one or more attribute name indexes that are the outputs of the keyword filter section 5. Take out one of the inputted attribute name indexes, and in the contents of the inputted text file Investigate whether a tag including the attribute name in this attribute name index exists, and when it exists It is a display 102 about a pair with the text file name inputted as the value of the attribute value tag corresponding to the attribute name tag. It leads and processing in which nothing is outputted when it does not display and exist in a user is performed to each of all the inputted attribute name indexes. The contents output section 9 of a document is carrying out the role called in subroutine by the contents output-control section 10 of a document. In addition, instead of outputting the pair of the value of an attribute value tag, and a text file name, you may make it display the value of an attribute value tag, and the positional information of a text file, and may make it display the value of an attribute value tag, a text file name, and its positional information.

[0042] The contents output-control section 10 of a document considers the set of the text file name which is the output of the success text file name sorting section 8 as an own input as it is, and the text file term document 1 is referred to. The need is accepted in the contents of the text file under set of the inputted text file name, and it is the Internet 103. It leads and is WWW. Access a server and it takes out one at a time. It repeats performing passing the contents output section 9 of a document to all the text file names under input with the attribute name index generated in a text file name and the keyword filter section 5. That is, when three text file names which passed as an input are received, the contents output section 9 of a document will be called 3 times. In addition, the success text file name sorting section 8 is the Internet 103. It is the Internet 103 by the contents output-control section 10 of a document using the contents, when the contents of the text file which it led and was incorporated from the WWWW server are saved at the magnetic disk drive etc. The count of access can be reduced.

[0043] Drawing 2 and drawing 3 are text file retrieval equipment 100. It is the flow chart which shows the example of processing. Hereafter, actuation of the gestalt of this operation is explained.

[0044] The keyword extraction section 4 is an input unit 101. If it leads and the retrieval input statement by natural language expression is received from a user (step S1), the attribute name storing dictionary 2 and the attribute value storing dictionary 3 will be referred to. Investigate sequentially from the head of a retrieval input statement, and when it is, whether there is any expression registered as a natural language expression in it in a retrieval input statement When it

is an attribute name, the attribute name index only containing an attribute name is outputted, and when it is attribute value, the attribute value index containing the pair of attribute value and a corresponding attribute name is outputted (step S2).

[0045] Next, the keyword filter section 5 inspects the list of the index outputted from the keyword extraction section 4, detects the part where the attribute name index and attribute value index of the same attribute are continuing, and deletes the attribute name index of the part (step S3).

[0046] Next, the success text file name sorting section 8 takes out the contents of the document of the text file name, and makes the contents Banking Inspection Department 7 of an integrated document judge delivery and success or failure paying attention to one text file name in the text file term document 1 with all the attribute value indexes outputted from the keyword filter section 5 (step S4).

[0047] The contents Banking Inspection Department 7 of an integrated document makes the contents Banking Inspection Department 6 of a document judge delivery and success or failure for this attribute value index and the contents of the text file paying attention to one of the attribute value indexes passed first, in order to inspect the passed contents of a document (step S5).

[0048] The attribute name tag with which the contents Banking Inspection Department 6 of a document has the attribute name included in the passed attribute value index in the contents of the passed text file exists. And the value of the attribute value tag which is the attribute name tag which existed and pair or [that inspect whether it is in agreement with the attribute value included in the passed attribute value index, and such an attribute name tag does not exist success in being in agreement] -- or even if it exists, when the attribute value is not in agreement, the contents Banking Inspection Department 7 of an integrated document is notified of a rejection (step S6) .

[0049] When, as for the contents Banking Inspection Department 7 of an integrated document, success is notified from the contents Banking Inspection Department 6 of a document (it is YES at step S7), It investigates whether it finish inspecting about all the attribute value indexes notified from the success text file name sorting section 8. When having finished inspecting still, attention is moved to one of NO) and the remaining attribute value indexes at the (step S8, and the contents Banking Inspection Department 6 of a document is made to judge delivery and success or failure for the attribute value index and contents of the text file (step S9). And when success is declared about all attribute value indexes in the contents Banking Inspection Department 7 of a document (it is YES at step S8), success is notified to the success text file name sorting section 8, the success text file name sorting section 8 uses the text file concerned as a success text file (step S10), and it progresses to step S11. On the other hand, when a rejection is notified from the contents Banking Inspection Department 6 of a document (it is NO at step S7), the contents Banking Inspection Department 7 of an integrated document notifies a rejection to the success text file name sorting section 8, and the success text file name sorting section 8 progresses to step S11.

[0050] After the yes-no decision about one text file finishes, when the unsettled text file remains into the text file term document 1 (it is YES at step S11), the success text file name sorting section 8 moves attention to one of text file names [them] (step S12), and makes the judgment of success or failure like a previous text file.

[0051] If it judges whether the success text file name sorting section 8 had at least one success file (step S13) and there is no one after finishing the yes-no decision to all the text files in the text file term document 1 (it is YES at step S11), the text file corresponding to the retrieval conditions inputted, for example will process displaying the purport.which one did not have on a user etc., and will end processing. On the other hand, when at least one success file exists, the text file name of all that success file and all the attribute name indexes outputted from the keyword filter section 5 are notified to the contents output-control section 10 of a document, and the contents output control of a document is made to start (step S14).

[0052] The contents output-control section 10 of a document takes out the contents of a document paying attention to one notified success file name, and the contents output section 9 of a document is made to start processing of delivery and the document concerned with all the notified attribute name indexes (step S15).

[0053] It investigates whether the contents output section 9 of a document has an attribute name tag with the attribute name of the attribute name index in a document paying attention to one notified attribute name index (step S16) (step S17), and if it is (it is YES at step S18), the value and the text file name concerned of the attribute value tag corresponding to the attribute name tag will be displayed on a display 102 (step S19). If there is nothing (it is NO at step S18), step S19 will be skipped. Next, if the contents output section 9 of a document investigated whether the unsettled attribute name index would remain on the notified attribute name index (step S20) and it remains in it, it will move attention to one of them (step S21), and will repeat the processing returned and mentioned above to step S17.

[0054] After finishing the processing about all attribute name indexes in which the contents output section 9 of a document was notified (it is NO at step S20), the contents output-control section 10 of a document When investigating

whether the unsettled thing remains to the text file notified from the success text file name sorting section 8 (step S22) and remaining in it, move attention to one of them, and the contents of a document of the text file name are taken out. The contents output section 9 of a document is passed, and is made to process with all the attribute name indexes notified from the success text file name sorting section 8 (step S23). It becomes processing termination after the processing about all success files finishes (it is YES at step S22).

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

EXAMPLE

[Example] Suppose that three text file names, "a file 1", "a file 2", and "a file 3", and URL of those are registered into the text file term document 1 so that it may illustrate to drawing 1 . Moreover, the contents of the file 1 show drawing 5 (a), and the contents of the file 2 presuppose that it is what is shown in drawing 5 (b). These files 1 and 2 are XML. It is described and the attribute and attribute value other than a text text are included. That is, the pair of the tag which the attribute name expressing the semantics written into the document of a <hotel> X hotel <<location> San Francisco </location> <price> $150 </price> <pet> possible < / pet> attached, and the value of the attribute is contained in the file 1. Similarly, a tag called a <hotel> Z hotel <<location> Seattle </location> <price> $180 </price> <pet> improper < / pet> is contained also in the file 2.
[0056] Moreover, the pair of an attribute name which is illustrated to drawing 1 , and its natural language expression shall be stored in the attribute name storing dictionary 2 in advance, and 3 groups of attribute value and a natural language expression which are illustrated to drawing 1 , and an attribute name shall be stored in the attribute value storing dictionary 3 in advance. In addition, if all prices are registered into the attribute value storing dictionary 3 with the actual value, since a number of registration will increase, you may make it register using a variable. That is, XXX If it registers as it is shown in the following table 3, when considering as the numeric value of arbitration, and the keyword extraction section 4 has a "dollar" behind the numeric value of arbitration, it will be the natural language expression XXX. Let what judged that a dollar existed and attached $ to the head of the actual value which existed be attribute value.
[Table 3]

ID=000005

[0057] Actuation of this example is explained to an example for the case where a user inputs the retrieval input statement by the following natural language by such premise.
Retrieval input-statement: "I want to find the information on the hotel which it is possible for a price to take a pet in 150 dollar, and is moreover near San Francisco"
[0058] The keyword extraction section 4 will change a retrieval input statement into a keyword train as follows with reference to the attribute name storing dictionary 2 and the attribute value storing dictionary 3, if the retrieval input statement from a user is received.
[0059] First, since the natural language expression "a price" of the head of a retrieval input statement exists in the attribute name storing dictionary 2, the attribute name "a price" registered into it and a pair by becoming is outputted as an attribute name index. Next, since a natural language expression "150 Dollar" exists in the attribute value storing dictionary 3, the pair of the attribute value "$150" and the attribute name "the price" which are registered in it and 3 groups is outputted as an attribute value index. Next, since a natural language expression "a pet" exists in the attribute name storing dictionary 2, the attribute name "a pet" registered into it and a pair by becoming is outputted as an attribute name index. Next, since a natural language expression "possible" exists in the attribute value storing dictionary 3, the pair of the attribute value "possible" and the attribute name "the pet" which are registered in it and 3 groups is outputted as an attribute value index. Next, since a natural language expression "San Francisco" exists in the attribute value storing dictionary 3, the pair of the attribute value "San Francisco" and the attribute name "the location" which are registered in it and 3 groups is outputted as an attribute value index. Next, since a "hotel" exists in the attribute name storing

dictionary 2, the attribute name "a hotel" registered into it and a pair by becoming is outputted as an attribute name index. Into a retrieval input statement, other natural language expressions which match the natural language expression registered into the attribute name storing dictionary 2 and the attribute value storing dictionary 3 cannot be found. Therefore, the following keyword trains are outputted sequentially from a top.

[0060] Attribute name index (attribute name "a price")
Attribute value index (attribute value "$150", attribute name "a price")
Attribute name index (attribute name "a pet")
Attribute value index (attribute value "possible", attribute name "a pet")
Attribute value index (attribute value " "San Francisco", attribute name "a location")
Attribute name index (attribute name "a hotel")

[0061] Next, the keyword filter section 5 unifies a redundant part with reference to the order of a list of an attribute name index and an attribute value index. Since it has ranked with next doors by the attribute name "a price" with same attribute name index (attribute name "price") and attribute value index (attribute value "$150", attribute name "price") in the case of the upper keyword train, an attribute name index (attribute name "a price") is deleted. Moreover, since it has ranked with next doors by the attribute name "a pet" with same attribute name index (attribute name "pet") and attribute value index (attribute value "possible", attribute name "pet"), an attribute name index (attribute name "a pet") is deleted. Since the attribute name index which should otherwise be deleted does not exist, finally the above-mentioned keyword train is compressed as follows.

[0062] (a) Attribute value index (attribute value "$150", attribute name "a price")
(b) Attribute value index (attribute value "possible", attribute name "a pet")
(c) Attribute value index (attribute value " "San Francisco", attribute name "a location")
(d) Attribute name index (attribute name "a hotel")

[0063] next, the success text file name sorting section 8 -- the contents of a document of the file 1 in the text file term document 1 -- the URL -- reliance -- the Internet 103 from the server which leads and corresponds -- acquiring -- the contents of a document, and the above-mentioned attribute value index (a) - (c) The contents Banking Inspection Department 7 of an integrated document is passed.

[0064] The contents Banking Inspection Department 7 of an integrated document is the contents of a document of a file 1, and one attribute value index (a). The contents Banking Inspection Department 6 of a document is passed.

[0065] The contents Banking Inspection Department 6 of a document is an attribute value index (a) in the contents of a document of a file 1. It investigates whether the tag of an inner attribute name "a price" exists. drawing 5 -- (-- a --) -- a file -- one -- a case -- corresponding -- a tag -- < -- a price -- > -- $ -- 150 -- < -- /-- a price -- > -- it is -- since -- the -- attribute value -- "-- $ -- 150 -- " -- having received -- attribute value -- an index -- (-- a --) It investigates whether it is in agreement with inner attribute value "$150." In the present example, since it is in agreement, success is returned to the contents Banking Inspection Department 7 of an integrated document.

[0066] The contents Banking Inspection Department 7 of an integrated document is the contents of a document of a file 1, and the following attribute value index (b). The contents Banking Inspection Department 6 of a document is passed.

[0067] The contents Banking Inspection Department 6 of a document is an attribute value index (b) in the contents of a document of a file 1. It investigates whether the tag of an inner attribute name "a pet" exists. drawing 5 -- (-- a --) -- a file -- one -- a case -- corresponding -- a tag -- < -- a pet -- > -- possible -- < -- /-- a pet -- > -- it is -- since -- the -- attribute value -- "-- possible -- " -- having received -- attribute value -- an index -- (-- b --) It investigates whether it is in agreement with inner attribute value "possible." In the present example, since it is in agreement, success is returned to the contents Banking Inspection Department 7 of an integrated document.

[0068] The contents Banking Inspection Department 7 of an integrated document is the contents of a document of a file 1, and the following attribute value index (c). The contents Banking Inspection Department 6 of a document is passed.

[0069] The contents Banking Inspection Department 6 of a document is an attribute value index (c) in the contents of a document of a file 1. It investigates whether the tag of an inner attribute name "a location" exists. drawing 5 -- (-- a --) -- a file -- one -- a case -- corresponding -- a tag -- < -- a location -- > -- San Francisco -- < -- /-- a location -- > -- it is -- since -- the -- attribute value -- "-- San Francisco -- " -- having received -- attribute value -- an index -- (-- c --) It investigates whether it is in agreement with inner attribute value "San Francisco." In the present example, since it is in agreement, success is returned to the contents Banking Inspection Department 7 of an integrated document.

[0070] Since the result of success was obtained by all attribute value indexes about the file 1, the contents Banking Inspection Department 7 of an integrated document notifies success to the success text file name sorting section 8, and the success text file name sorting section 8 considers a file 1 as a success file.

[0071] Next, the success text file name sorting section 8 is the Internet 103 to reliance about the URL in the contents of a

document of the file 2 stored in the text file term document 1. It leads, and incorporates from the corresponding server and success or failure is judged using the contents Banking Inspection Department 7 of an integrated document like the previous file 1. In this case, since a pet attribute and a location attribute are not satisfied, a file 2 serves as a rejection. Similarly, the judgment of success or failure is performed also about the remaining file 3. Here, a file 3 is also judged to be a rejection and a success file presupposes that it was it only a file 1.

[0072] Next, the success text file name sorting section 8 is an attribute name index (d) about a file name 1 as a success file name. The contents output-control section 10 of a document is passed.

[0073] The contents output-control section 10 of a document acquires URL of a file name 1 from the text file term document 1, and is the Internet 103 to reliance about it. The upper server is accessed, the contents of a document of a file name 1 are acquired, and it is an attribute name index (d). The contents output section 9 of a document is passed.

[0074] The contents output section 9 of a document is an attribute name index (d) in the contents of a document of a file 1. It investigates whether an attribute tag with an attribute name "a hotel" exists. since the corresponding tag <hotel> X hotel <a /hotel> exists in the case of the file 1 of drawing 5 (a) -- the attribute value "X hotel" -- taking out -- a file name 1 -- display 102 It displays.

[0075] Drawing 4 is the block diagram showing the example of a hardware configuration of the text file retrieval equipment of this invention. The text file retrieval equipment of this example A processor 200 (CPU), The auxiliary storage units 201, such as a magnetic disk drive, and the interface 202 of those, The memory 203, such as RAM, and the interface 204 between the Internet 103, The record medium 205 which can machine read CD-ROM, semiconductor memory, etc., It consists of buses 209 which connect between the interface 206, an input unit 101, its interface 207, a display 102, its interface 208, and CPU200, memory 203 and an interface 202,204,206-208.

[0076] The program for text file retrieval is recorded on the record medium 205, and the text file term document 1, the attribute name storing dictionary 2, and the attribute value storing dictionary 3 of drawing 1 are loaded on memory 203 or an auxiliary storage unit 201 by installing this program through an interface 206. Moreover, this program operates CPU200 by controlling actuation of CPU200 as the keyword extraction section 4 of drawing 1 , the keyword filter section 5, the contents Banking Inspection Department 6 of a document, the contents Banking Inspection Department 7 of an integrated document, the success text file name sorting section 8, the contents output section 9 of a document, and the contents output-control section 10 of a document.

[0077] The gestalt of the above operation is WWW about this invention. Although applied to retrieval of the upper homepage, it is applicable also to the retrieval to a database like the technique indicated to JP,5-67136,A.

[Translation done.]

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the block diagram of the text file retrieval equipment of the gestalt of operation of this invention.
[Drawing 2] It is the flow chart which shows the example of text file retrieval equipment processing of the gestalt of operation of this invention.
[Drawing 3] It is the flow chart which shows the example of text file retrieval equipment processing of the gestalt of operation of this invention.
[Drawing 4] It is the block diagram showing the example of a hardware configuration of the text file retrieval equipment of this invention.
[Drawing 5] XML It is drawing showing the example of description of the used text file.
[Drawing 6] It is drawing showing the example of description of HTML.
[Drawing 7] It is drawing showing the example which displayed the example of description of HTML of drawing 6 by the browser.
[Description of Notations]
1 -- a text file term document and 2 -- an attribute name storing dictionary and 3 -- an attribute value storing dictionary and 4 -- the keyword extraction section and 5 -- the keyword filter section and 6 -- the contents Banking Inspection Department of a document, and 7 -- the contents Banking Inspection Department of an integrated document, and 8 -- the success text file name sorting section and 9 -- the contents output section of a document, and 10 -- the contents output-control section of a document -- it comes out.

[Translation done.]